VOLUME 55
NUMBER 3

Whole No. 251
1943

Psychological Monographs

JOHN F. DASHIELL, *Editor*

The Application of the Theory of Physical Measurement to the Measurement of Psychological Magnitudes, with Three Experimental Examples

By

THOMAS WHELAN REESE, PH.D.

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Publications Office
NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

ACKNOWLEDGEMENTS

The author wishes here to recognize his deep indebtedness to Professor H. E. Garrett and Dr. John Volkmann who have made this study possible.

Cette échelle permet, non pas à proprement parler la mesure de l'intelligence,—car les qualités intellectuelles ne se mesurent pas comme des longeurs, elles ne sont pas superposables,— . . . !

A. Binet et Th. Simon

CONTENTS

THE APPLICATION OF THE THEORY OF PHYSICAL MEASUREMENT TO THE MEASUREMENT OF PSYCHOLOGICAL MAGNITUDES, WITH THREE EXPERIMENTAL EXAMPLES

PART I

INTRODUCTION

THE PROBLEM

When an observer in a discrimination experiment makes a differential response to a stimulus, he is responding with respect to some *discriminable characteristic.* This characteristic may be defined in general terms as the combined effect upon discriminatory behavior of the experimenter's operations of stimulation and instruction. Discriminable characteristics, so defined, are "subjective" only in the sense that they are clearly distinguished from the stimulus-correlates. They are not identified with a private, immediate experience.

Every stimulus to which the observer can make a verbal response must have given rise to at least one discriminable characteristic. However it is probable that every stimulus produces more than one. For example, an auditory stimulus produces such discriminable characteristics as pitch and loudness; a visual stimulus produces such characteristics as hue, brilliance and saturation.

With respect to many discriminable characteristics the observer will be able to make verbal responses of "greater than," "less than," "not greater than" and "not less than." Those discriminable characteristics to which this type of verbal response can be made may be said to exist in discriminable degrees.

The purpose of this study is to examine the possibility of measuring those discriminable characteristics that exist in discriminable degrees.

Many discriminable characteristics have known physical correlates. For example, the chief physical correlate of the discriminable characteristic of weight is the "physical weight" of the object. The measurement of "physical weight," in fact the measurement of all physical correlates, is a problem for the physicist, but the measurement of the so-called subjective magnitudes is a problem for the psychologist. This statement assumes, of course, that there is a difference between physical and subjective magnitudes. Although this difference seems obvious from the common sense point of view, it is not too easy to frame a monistic operational definition of the two types of magnitudes.

A discriminable characteristic has been defined as the combined effect on discriminatory behavior of the experimenter's operations of stimulation and instruction. The instructions include the directions to make a judgment of "greater than," "less than," "not greater than" and "not less than." The discriminatory behavior includes four different responses, one for each of these judgments. If for given conditions of stimulation and under the appropriate instructions, the observer makes these responses so as to meet some arbitrarily predetermined standard of consistency, it is said that he is able to discriminate degrees of the characteristic. But this discriminatory response is fundamental to both physical and subjective magnitudes. The two can-

not be differentiated in terms of this basic response.

If there is a difference it must be found in differences in the operations of stimulation or in the remaining operations of instruction.

There does seem to be one difference between the procedure adopted by the physicist and that adopted by the psychologist. Having identified a characteristic, the physicist in the interest of consistency and greater discriminatory power, usually abandons it for another characteristic which is correlated with the first one. For example, the characteristic of subjective weight may be identified by a series of operations which involve, among others, the operation of hefting. The physicist will find that he is able to construct a magnitude that correlates with the original subjective magnitude but which substitutes operations involving balances for the operation of hefting. Furthermore the discriminable characteristic is changed from subjective weight to some spatial characteristic such as the position of a pointer.

In many cases the change of characteristic will be a good deal less obvious than it was in the case above. The difference between the operations for scaling subjective and physical length is a case in point. Even here examination shows that the operations adopted by the physicist entail a change of discriminable characteristic. The scaling of subjective length will entail a judgment of the overall length of the stimuli. But the physicist will place the two stimuli side by side and give instructions that demand that the observer disregard the overall length and make a judgment concerning the presence or absence of a difference.

By this means the physicist is able to extend the scale beyond those limits imposed by the low discriminatory capacity of the observer with respect to the original discriminable characteristic. He is able to define magnitudes above the upper limit for the original characteristic and below its absolute threshold. He is also able to increase differential sensitivity.

Although the psychologist may change his instructions and alter the conditions of stimulation in the interest of consistency and greater discriminatory capacity, unlike the physicist, he cannot change the discriminable characteristic.

There are not only a large number of discriminable characteristics, but also a large number of ways in which these characteristics may differ from one another. But characteristics not only differ from another; many characteristics are similar to one another in respect of certain aspects they have in common. For example, hue, brilliance and saturation are all mediated by the same sensory mechanism. These features, by means of which discriminable characteristics may be described and classified, will be called, for the purpose of this discussion, the *aspects* of discriminable characteristics. Certain discriminable characteristics may be classified as qualitative and others as intensitive (Boring, 4); quality and intensity may be said to be aspects of discriminable characteristics. Auditory pitch has an aspect of quality and auditory loudness an aspect of intensity.

It may be possible to find an aspect that makes it difficult to measure all the discriminable characteristics possessing it, regardless of any other aspects which the characteristics may or may not have in common. Likewise, aspects may be found which permit the characteristic to be measured easily.

Now, it is obvious that it is impossible to *prove* that *every* discriminable characteristic is measurable unless every char-

acteristic is actually measured successfully. Only a probable inference can be drawn from partial evidence. The validity of the inference will depend, in great part, upon the adequacy of the sample. Adequacy may refer simply to the number of cases that are selected at random. By this definition an *adequate* sample is one that contains the number of cases that allows reasonable probability that the distribution of the aspects is approximately the same in the sample as in the population being sampled. There is no criterion for adequacy defined in this way. All one is able to say is that the greater the number of cases, the greater is the probability that the sample is adequate.

For the purpose of this study there is another way to look at adequacy. If the presence of a certain aspect leads to the belief that the characteristic having it will be difficult to measure, then it is possible purposely to select those discriminable characteristics which have those aspects that offer the least chance for successful measurement. If, then, they are measured successfully, it may be argued that the probability for the successful measurement of *all* discriminable characteristics is increased. The validity of this procedure may be further increased by selecting several *different* difficult aspects. By this method of selection of the characteristics to be measured, both conceptions of an adequate sample are combined.

The first thing, then, that needs to be done is to classify the aspects of discriminable characteristics and to choose several of them according to the above principles.

There are at least four categories of classification that might be important:

1) The sense modality.

2) The aspects of "conscious dimensions" as described by Boring (4): a) the qualitative dimension as exemplified by auditory pitch; b) the intensitive dimension as exemplified by loudness; c) the extensitive dimension as exemplified by length or volume and d) the protensitive dimension as exemplified by time.

3) The characteristics may have the aspect of being palpable or impalpable. *Impalpable* is Titchener's translation of *unanschaulich,* the adjective given to Ach's *Bewusstheit* (awareness). The characteristic may be a "vague, intangible conscious content that is not image or sensation," to quote Boring's (3) description of *Bewusstheit.*

The difference between palpable and impalpable may be exemplified by comparing auditory pitch and the experienced difficulty of doing a mental test item.

The pitch characteristic of a tone is not impalpable. Even if it fades and becomes intangible and vague, the nature of the experience is such that it may be recaptured on a second presentation of the stimulus. But the experienced difficulty of doing a mental test item is of an entirely different order. Here the original impression may be relatively impalpable and when once lost can never be regained, for it is extremely unlikely that the experienced difficulty will be the same on a second presentation of the same item.

4) Stimulus correlation. The discriminable characteristic may be known to be correlated with one or several aspects of the stimulus, as, for example, pitch is correlated with frequency and intensity, loudness with intensity and frequency. On the other hand the correlated stimulus aspect may be unknown or exceedingly complex.

Several subjective scales have been constructed or are in the process of construction. Stevens (38) has constructed a scale for loudness. Stevens, Volkmann and

Newman (39) have constructed a scale for pitch, which later has been revised by Stevens and Volkmann (42). Taves (44) has constructed a scale for visual numerousness (perceived number) and Taback (43) for perceived weight.

Thus there are magnitude functions that may be classified under three sense modalities, vision (numerousness), audition (pitch and loudness) and kinaesthesis (weight). They may be classified under three conscious dimensions, qualitative (pitch), intensitive (loudness and weight) and extensitive (visual numerousness). The characteristics are relatively palpable and have known stimulus correlates. The stimulus correlates of pitch are frequency and intensity, of loudness are intensity and frequency; while the correlate of numerousness is the actual number of stimulus objects, and the correlate of perceived weight is the physical weight.

The three discriminable characteristics with which this experiment will deal are: 1) visual rate, or the perceived rate of the flash of a lamp, 2) the experienced difficulty of items in a memory span test (digits) and 3) the experienced difficulty of multiple choice items in a vocabulary test.

It will be seen that these characteristics differ not only from those that have already been scaled but differ radically from each other. The chief differences are in respect of palpability and stimulus correlation. Visual rate has the aspect of palpability and of a known stimulus correlate, i.e., the actual rate of the flash of the lamp. The experienced difficulty of items in a memory span test is relatively impalpable and the important stimulus correlate is known to be the number of digits in the series. In the difficulty of words in the multiple choice vocabulary test the characteristic is impalpable and the stimulus correlate is unknown. It is, for example, impossible to say that it is the number of letters in the word or its length in inches.

The "physical dimensions" of these characteristics would seem to be different. Visual rate is protensitive and although it would be difficult to fit the physical dimension associated with experienced difficulty into Boring's classification, at least it does not seem to be solely protensitive.

The important differences for this study are those of palpability and stimulus correlation. Impalpability presents some difficulties, as a characteristic that is intangible and fleeting will be more difficult for the subject to judge than one that is not, and the impossibility of recovering the experience will present some technical difficulties in designing the experiment. The lack of a stimulus correlate will also present some difficulty. In fact Guilford (21) has said that the complete psychophysical treatment of mental test data is impossible because of the lack of a physical evaluation of the stimulus.

Guilford emphasizes the importance of this problem of the physical correlate because it seems to be the chief factor in obscuring the common ground between psychophysical and mental test problems (22). He has found the relation between the psychological difficulty of a test item and a corresponding physical evaluation of the items, using as his definition of difficulty "percentage failing" and using as his test items certain of the Seashore tests.

There are in reality two problems here. One is the measurement of the discriminable characteristic and the other is the relation between the physical and

the subjective magnitudes. So far as *measurement* is concerned the problem of the stimulus correlate solves itself. It will be shown later in this study that a physical evaluation of a correlated stimulus variable is unnecessary for the measurement of a subjective magnitude, provided only that one has some means by which stimuli of different subjective magnitudes may be physically identified. So far as the relation between the two variables is concerned, naturally it will be impossible to demonstrate a relation between some variable and a stimulus correlate if the stimulus correlate cannot itself be identified or measured.

PART II

THE LOGIC OF MENTAL MEASUREMENT

SECTION A. INTRODUCTORY

RECENTLY a number of physicists (and logicians) together with a group of psychologists have leveled a particularly vigorous attack against the theoretical concepts upon which psychologists have based their practice of measurement. The criticism has come to a head with the recently published *Final Report* of the *Committee appointed to consider and report upon the possibility of Quantitative Estimates of Sensory Events* (14). The members of this committee were drawn from Sections A (Physics) and J (Psychology) of the *British Association for the Advancement of Science.*

In the following sections the criticisms of both the physicists and the psychologists will be examined in some detail. Suffice it to say here that the physicists have claimed that measurement in any true sense is impossible in psychology. They base this conclusion on what they consider to be the fact that none of the attempts at measurement in psychology meet the necessary logical requirements for fundamental measurement.

They argue that psychologists must then do one of two things. They must either say that the logical requirements for measurement in physics, as laid down by the logicians and other experts in the field of measurement, do not hold for psychology, and then develop other principles that are logically sound; or they must admit that their attempts at measurement do not meet the criteria and both cease calling these manipulations by the word "measurement" and stop treating the results obtained as if they were the products of true measurement. For example Guild, who seems to have taken the most extreme position against the possibility of measurement in psychology, says, "To insist on calling these other processes[1] measurement adds nothing to their actual significance but merely debases the coinage of verbal intercourse. Measurement is not a term with some mysterious inherent meaning, part of which may be overlooked by the physicists and may be in course of discovery by psychologists. It is merely a word conventionally employed to denote certain ideas. To use it to denote other ideas does not broaden its meaning but destroys it: we cease to know what is to be understood by the term when we encounter it; our pockets have been picked of a useful coin" (13).

The *Final Report* of the committee of the British Association for the Advancement of Science holds out hope for a third solution, when in paragraph 10 it states, "Some members, perhaps all, admit that their opinion might change if new facts were established; but the facts that would be necessary for this purpose are not of the kind that can be established by any experimental method at present in general use" (14).

For convenience the discussion will be divided into sections. The logical criteria which the physicists claim that all measurement must meet will be presented and discussed in Section B; the practical operations necessary for the fulfillment of these criteria will be discussed in Section C; the position of Stevens, who has given the most vigorous reply

[1] I.e., the method of equal appearing intervals and the method of just noticeable differences.

to the logicians' criticisms, will be discussed in Section D; the specific criticisms of the psychological operations for measurement which have been raised by the physicists will be presented and discussed in Section E; the criticisms of psychologists of their own methods will be presented and discussed in Section F; the special problem of zero subjective magnitudes will be discussed in Section G; it will be shown in Section H that measurement in psychology does not depend on the prior measurement of any other magnitude; Section J contains a brief summary of the discussion up to that point.

SECTION B. THE LOGICAL REQUIREMENTS OF MEASUREMENT[2]

A distinction must be made at the outset between measurement defined as the construction of a scale and measurement defined as the use of the scale after it has been constructed. Utter confusion will result from the confounding of these two definitions. The use of a measuring scale after it has been constructed is a more or less simple matter involving the comparison of the object to be measured with the standard scale. The word measurement is *never* used in this sense in this paper. As here used it refers to the more fundamental problem of scale construction.

Measurement, according to Campbell, is the assignment of numerals to systems[3] according to scientific laws. The scientific laws spring from the relations demonstrated between the systems with respect to a certain magnitude.

The first requirement for measurement is that it must be possible to arrange the systems to be measured in respect of a given magnitude, in an *order,* with respect to that magnitude. The result of this operation is known as an ordinal scale. To do this it must first be demonstrated by some operations that the relation between the systems is *transitive* and *asymmetrical.*

If the symbol $\rightarrow$ means "bears a certain relation to" and $\not\rightarrow$ means "does not bear that relation to," it must then be shown experimentally that the relation in question is asymmetrical, that is, if

$$a \rightarrow B \text{ then } B \not\rightarrow A \qquad \text{I (I)}$$

and transitive, that is, if

$$A \rightarrow B \text{ and } B \rightarrow C, \text{ then } A \rightarrow C. \qquad \text{II (I)}$$

If the above symbols are replaced by $>$ (greater than) and $\ngtr$ (not greater than) or by the converse $<$ (less than) and $\nless$ (not less than), it would be necessary to show that, if

$$A > B \text{ then } B \ngtr A \qquad \text{I (II)}$$

and that if

$$A > B \text{ and } B > C, \text{ then } A > C. \qquad \text{II (II)}$$

It will be seen that the relation $=$ does not exist in such a series. An example of such a series without the relation $=$, given by Campbell, is the direct line of male descent. The generating relations are "ancestor of" and "descendant of."

The relation $=$, as between A and B, is associated with the following propositions:

$A = B$ if, and only if,

$$1)\ A \ngtr B \text{ and } A \nless B \qquad \text{III}$$
$$2)\ \text{if } A > C \text{ then } B > C \qquad \text{IV}$$
$$3)\ \text{if } A < C \text{ then } B < C. \qquad \text{V}$$

The relation $=$ defined in this manner is always transitive, that is, if

$$A = B \text{ and } B = C \text{ then } A = C \qquad \text{VI}$$

and is always symmetrical, that is, if

$$A = B \text{ then } B = A. \qquad \text{VII}$$

[2] In this section the author has borrowed liberally from Campbell (5, 6, 7), Guild (13) and Cohen and Nagel (9). No references are given except in those cases where an author is quoted directly or an illustrative example employed by the author is used in the text.

[3] Campbell apparently uses the term "system" to mean any objects with which the physicist deals. The term would seem to include anything from pieces of wood or rock to electric lamps and voltage dividers.

On casual inspection it might seem difficult for a given relation to satisfy III and yet fail to satisfy IV and V but if the example of Campbell's, given above, is examined it will be seen that it would be possible for A to be neither the ancestor nor the descendant of B and thus satisfy III, and yet A cannot then satisfy IV or V. To quote Campbell ". . . no two males can have all the same ancestors and descendants" (6).

According to Campbell a magnitude must have the relation =, for, as will be seen later, this relation is necessary in the construction of an additive scale.

He sums up the first conditions for measurement as follows: "The first condition of measurement, namely that a magnitude must be capable of order, can now be stated formally as follows. The systems measured must, in virtue of the property concerned, be a field of a pair of converse T.A.[4] relations and the T.S.[5] relation associated with them; every system must be $>$ or $<$ or $=$ every other, and must be $=$ at least one other. The first law of measurement is the statement that this condition is fulfilled" (6).

The rule for assigning numerals to represent a series in which the above relations have been established is: if $A > B$ then the numeral assigned to A must be greater than the numeral assigned to B;[6] conversely, if $B < A$, then the numeral assigned to B must be less than the numeral assigned to A. If $A = B$, then the numeral assigned to A must be the same as the numeral assigned to B.

According to Campbell, the existence of the relation = is one of the things that distinguishes the order characteristic of magnitudes from that which is characteristic of numerals. Numerals, by which is meant simply a group of conventional signs or marks on a piece of paper, obtain their order by convention. The order is not determined by facts such as the order existing in the family tree. If only one of the many numeral series is used, every member is either greater or less than every other member. There is no relation =. (Naturally if several series were combined, such as the decimal and fraction series, than it would be possible to find two that were equal to each other, as $1.5 = 1\frac{1}{2}$.)

However the most important difference between numerals and magnitudes is that the order of the numerals is conventional while the order of the systems in respect of the magnitude is determined by experimental operations.

Numerals have by convention a transitive, asymmetrical relation. Now if they are going to be used to represent the order of the systems in respect of a certain magnitude, it must be *shown experimentally* that the relation between the systems which they represent is also transitive and asymmetrical. If it is impossible to show this then the numerals that have been assigned are meaningless in as much as the conventional relations between them do not express the relations between the systems.

There is nothing in the experimentally established relation $A > B$ that tells *what* numeral is assigned to A. The rule simply states that it must be greater than that assigned to B. As yet there are no operations to determine by how much $A > B$, so the assigned numerals cannot reflect a relation that has not been established. In other words, if 2 is assigned to B and 4 to A, it is impossible to say that A is twice as great as B because it has not yet been shown experimentally

[4] Transitive asymmetrical.
[5] Transitive symmetrical.
[6] The problem of how a numeral, which has been defined as a symbol, can be greater than another numeral is taken up later.

that A *is* twice B. In other words, *the numerals can express only those relations that have been shown experimentally to exist between the systems to which they are assigned.*

An interesting example of an ordinal scale is the Mohs scale of hardness. Mohs, in developing his scale for the hardness of minerals, tried to ordinalize minerals according to the relation "scratches." The operation by which the relation of the minerals was to be determined was to attempt to scratch one mineral with another. He selected ten minerals to represent particular points on the scale and assigned numerals to them. The numerals ranged from 1 to 10, where the numeral 1 was assigned to that mineral which could be scratched by every other mineral and which could scratch no mineral; and 10 was assigned to that mineral which could scratch every other one and be scratched by none. It should be noted that there is nothing in the operations adopted by Mohs that tells how many more times as hard one mineral is than another. It simply tells that one mineral is harder than another, as defined by the operation of scratching, and therefore should be assigned a higher numeral.

Later attempts at measuring hardness, *defined by other operations,* such as microscopic measurement of the depth of a scratch made by a diamond under constant pressure or the amount of work done in grinding away a certain weight or volume of material, have shown that the interval between Mohs' hardness of 10 and 9 was greater than the interval between 9 and 1 (37).

Mohs assumed that his relation of "scratches" was transitive and asymmetrical and that the = associated with it was transitive and symmetrical. This has been shown to be false. Some minerals have been found that satisfy III but not IV or V,[7] that is they cannot scratch each other, and yet have different powers of scratching a third mineral. Because of this, according to Campbell, hardness as defined by the operation of scratching is not a magnitude at all.

In order to determine *what* numeral should be assigned to A if $A > B$ or, in other words, in order to be able to construct an additive or extensive scale, the property being scaled must be capable of being "added." The second requirement for measurement, then is that it must be possible to find some operation by which the magnitudes of two = systems may be combined to form systems that are $\neq$. In order to be additive the proposed method of combination must meet the following conditions: if

$$A = A' \text{ and } B > 0, \text{ then } A + B > A' \qquad \text{VIII}$$
$$A + B = X \text{ then } B + A = X \qquad \text{IX}$$
$$A = A' \text{ and } B = B' \text{ then } A + B = A' + B' \qquad \text{X}$$

and

$$(A + B) + C = A' + (B' + C'). \qquad \text{XI}$$

To quote Campbell again, "The statement that these conditions are fulfilled by any proposed method of addition defined by + and ( ), applied to systems possessing any magnitude defined by >, <, and =, is the second law of measurement of that magnitude" (6).

It is now necessary to have a rule for the assignment of numerals to the systems to represent the new relations that have been obtained by the operation of addition.

In discussing the assignment of numerals it is well to stress again that it is *numerals* that are assigned and not numbers. As Campbell says, ". . . it would be difficult to avoid the impression that the conception of number and the rules of

[7] Throughout this monograph roman numerals will be used to refer to the logical requirements presented and discussed in this section.

arithmetic were concerned in the matter. Actually they are not concerned. Of course, they are closely connected with measurement; but if we fail to recognize that they are not essential we shall not understand the connection" (6).

The first operation for the construction of an additive or extensive scale is to select some system that belongs to the series of the magnitude. The selection is entirely arbitrary. Then another system is found that is = to this first system. If the numeral A is assigned to the system first selected, then the numeral A′ is assigned to that system that is = to A. Since the systems must be identifiable, that is since it is absolutely necessary that they can be told apart by some method, it will be convenient to call the second system A′ to indicate that it is of the same magnitude as the system A but is a different system. The ′ is not to be taken as an indication of a different numeral but of the same numeral assigned to a different system.

The next step is to "add" the systems A and A′ and seek a system that is = to their combination. To this system the numeral B can be assigned. Another system that is = to B is then found and assigned the numeral B′, B and B′ are "added" and another system C is found such that B + B′ = C, and so on.

Using the ordinary numeral series instead of the alphabet and arbitrarily assigning 1 to the systems previously assigned A, B would equal 1 + 1 or 2. Likewise C = 2 + 2 or 4. Naturally numerals may be assigned to intermediate systems. If a system A″ is found that = A′ = A, then it is theoretically possible to find a system = A + A′ + A″ to which the numeral 3 would be assigned.

It is easily seen that a great advance has been made when it is possible to construct an additive scale. The ordinal scale did not tell by how much A > B because there was nothing in the relations established by experiment that determined by how much A > B. But now, once having chosen a standard, all the other magnitudes are uniquely determined. It is known for example that C is not only > B but also that C = B + B′, because the operations performed on the systems have determined this relation experimentally.

It is now time to examine more closely the rules for the assignment of numerals. Before this is done it will be necessary to clear up some questions of terminology. Much of the difficulty of the subject of measurement seems to stem not only from the confusion of the meanings of the words *number* and numeral, but also from the fact that the word *number* itself has several meanings. The following definitions have been adopted in this paper.

Numeral. A numeral is a sign or symbol that may be conventionally used to represent a number. In other words it is simply a black mark on a piece of paper. There are several conventional numeral series, 1, 2, 3, etc., or A, B, C, etc.

Number. Russell's definition of *number* is not used in this paper. Number is here regarded as a discriminable characteristic of systems that may be measured as any other discriminable characteristic. The term is used in much the same way as Stevens (40) has used the term "numerosity." "Numerosity," he says, "is a property defined by certain operations performed upon groups of objects." In his discussion he begins by saying that it is possible to establish a rank order of groups of objects [beans for example] in respect of numerousness[8] simply by look-

[8] This kind of numerousness is called subjective number in this paper.

ing at the piles of beans and judging which of the piles is largest, etc.

"We know from experience, however, that greater reliability can be had if we rank-order the groups by pairing successively one bean from each group until one group is exhausted. Then if any beans remain in the other group, that group is said to have the greater *numerosity*. . . . If the pairing exhausts both groups simultaneously, their numerosity is equal . . .," etc. It can be seen that number defined in this way can be measured fundamentally (see also Campbell, 6).

As this is not the usual definition of number it has been decided to call this kind of number "objective number" to distinguish it from number as a logical concept as used by Russell and from subjective number (numerousness).

It will be remembered that systems arranged in an experimentally established order had numerals assigned to them by the following rule: If A >B, then the numeral assigned to A must be greater than the numeral assigned to B and conversely if B <A then the numeral assigned to B must be less than the numeral assigned to A. Furthermore if A = B, the numeral assigned to A must be equal the numeral assigned to B.

It is now possible to ask the question how can a numeral, which has been defined as a mere sign or symbol, have a magnitude? In short how can one numeral be > or < or = any other numeral? One possible answer has already been implied, when it was assumed that the numerals used were those that are conventionally arranged in an order. In other words >, with respect to numerals, means "following" in the numeral series. Thus 2 > 1 because it follows 1 and D > B because it follows B in the numeral series.

However it is extremely important to note that it is not necessary to use a conventional numeral series in the construction of the scale. Any other group of numerals would do as well. In the event that a group of numerals which did not have a conventional order were used in the construction of the scale, the numerals would be arbitrarily assigned to the various systems in the experimentally established ordinal series. But once having been assigned, their order, so far as measurement of the particular magnitude is concerned, would be uniquely determined by the order of the systems to which they were assigned. In other words, though the numerals did not originally possess an order, once they have been assigned to a group of systems, they represent the relations that have been determined experimentally between the systems in the ordered series. If the relation between the systems is transitive and asymmetrical, then the numerals express this transitive and asymmetrical relation; if the relation between the systems is intransitive and symmetrical, the numerals express an intransitive symmetrical relation. Whatever relations the numerals express, they express only by virtue of the fact that these relations were shown to exist between the systems to which they have been arbitrarily assigned.

It should be noted that this statement is also true if numerals with a conventional order are used. It so happens that by convention such a series as 1, 2, 3, etc., is transitive and asymmetrical and it so happens that the relation that must first be established between the systems is also transitive and asymmetrical. In other words both are an ordered series, one is ordered by convention and the other by experimental operations. For convenience, then, we assign numerals to the systems so that if A > B, the

numeral assigned to A is greater than the numeral assigned to B, always remembering that the relation "greater than" when applied to numerals is a matter of convention.

There is nothing in the size of the numeral assigned to A that makes A greater than B. For example if $A > B$ and the numeral 2 is assigned to B, then any numeral that is conventionally greater than 2 may be assigned to A, say 4. But suppose that $A > B$ and the numeral 2 is arbitrarily assigned to B and the numeral 1 is assigned to A. The fact that 1 is assigned to A does not now make A less than B. In fact it works the other way; the fact that A has been shown to be greater than B means that the numeral 1 assigned to A *must* be interpreted as greater than the numeral 2 assigned to B. A new series is brought into being by these operations in which the *numeral* 1 is "greater than" the numeral 2. This means that, that so far as this particular magnitude is concerned, the numeral series 1, 2, 3, etc., cannot be used or interpreted in the usual way, i.e., $2 > 1$, etc.

The seeming absurdity of $1 > 2$ arises because one is accustomed to think of 1 and 2 as numbers, not numerals. The number conventionally represented by the numeral 1 is certainly not greater than the number conventionally represented by the numeral 2, but the numeral 1 may be regarded as either greater or less than the numeral 2 depending on the magnitudes of the systems to which these numerals are assigned. This would be highly inconvenient and it is much more reasonable to use the conventionally ordered series. But it is also important to see that it is not necessary to use the conventional series, in order to make clear the fact that the numerals add nothing to the experimentally determined relations.

The same reasoning that applies to the use of nonconventional numerals to represent systems in an ordered series applies to their use for representing the systems in an additive series. For example it was said that if $B = A + A'$ and the numeral 1 is assigned to A, then the numeral $1 + 1$ or 2 would be assigned to B. But what meaning can be assigned to the statement, "the *numeral* 1 + the *numeral* 1 = the *numeral* 2"? There is nothing about the numerals 1 and 2 *qua* numerals that would justify the conclusion that $1 + 1 = 2$. However as convention numerals, $1 + 1 = 2$ because the numeral 1 has been conventionally assigned to a certain objective number and the numeral 2 has been assigned to another objective number. Furthermore if we call the number to which the numeral 1 is assigned X and the number to which 2 has been assigned Y, and it is possible to show by a series of operations involving addition that $Y = X + X'$, then it is possible to state that $1 + 1 = 2$. But it must be carefully borne in mind that so far as measurement is concerned this *numerical* statement has no meaning apart from the experimentally established relations between the objective numbers X and Y. When the criteria for measurement have been met the relations between the systems are analogous to the relations between objective numbers which are represented by the ordinary numeral series 1, 2, 3, etc. It is then possible to use the ordinary numeral series in its conventional sense and apply the powerful tool of arithmetic to the symbols with the knowledge that these arithmetic manipulations represent, with only a small margin of error, the actual physical operations that might be performed on the systems themselves.

But it is not necessary to use the conventional numeral series, it is only convenient to do so. Any other numeral series, or any other group of numerals could be used, though it would be necessary to construct a new arithmetic, i.e., new laws for the manipulation of numerals, if it was necessary to manipulate these numerals rather than perform actual operations on the systems.

The procedure outlined above for the construction of an additive scale results in what Campbell names an A-magnitude, also sometimes called a fundamental magnitude. An A-magnitude (or fundamental magnitude) is one for which a practical operation of addition may be found. There is another larger and very important group of magnitudes which are called by Campbell, B-magnitudes, also sometimes called derived magnitudes. A B-magnitude (or derived magnitude) is one that is measured in terms of an A-magnitude. It cannot be measured directly because it is impossible to find a practical operation for addition that will meet VIII, IX, X, XI. B-magnitudes may be of two kinds. To quote Campbell (6), "The property measured in this manner may be nothing but that of being subject to the numerical law, and may be indefinable apart from that law. But it has often happened that the discovery of a numerical law, and of the constants associated with it, has enabled us to measure in this way a property that had previously been suspected of being a magnitude, but had not been actually measured."

The example usually given to illustrate B-magnitudes is *density*. Density = mass/volume. Both mass and volume are A-magnitudes. As Campbell shows, it is suspected that density is a magnitude because liquids might be arranged in an order of magnitude by defining "denser than" by *floats on*. It could be shown that this relation is transitive by showing that if A floats on B and B floats on C, then A will float on C. It could also be shown that the relation is asymmetrical. If A floats on B, then B does not float on A. Equality could be defined as that state in which neither liquid will float on the other permanently.

If liquids are then arranged in an order defined by the quotient mass/volume and this order is identical with the order obtained by defining density by flotation, it is possible to say that the property measured by the quotient mass/volume is the same as the property measured by flotation. To quote Campbell, ". . . the discovery of the law[9] has enabled us to measure a property previously immeasurable"(6).

It may be asked why the quotient mass/volume is thought to measure the same magnitude that is measured by flotation. Campbell lays down the general principle that "the conception of a magnitude is inseparable from that of the order characteristic of it. It is natural, therefore, to regard as the same magnitudes, or as *magnitudes of the same kind,* properties that invariably have the same order"(6).

Temperature is often mentioned as an example of a magnitude which is measured without a practical operation for addition. Guild (13) defines temperature as "the condition of an object in virtue of which it may feel hot or cold to the touch. . . ." He states further that "Experiment has shown many observable relations of a general kind between the temperature of bodies and their measurable properties. The length and electrical resistance of a given rod,

[9] The "law" refers to the fact that mass/volume is a constant for given liquids under defined conditions.

for example, are usually greater when the rod feels hot than when it feels cold."

In order to measure temperature as a B-magnitude it is only necessary to choose some measurable property which varies continuously with it. This having been done it is possible to postulate a law relating the property chosen to temperature. For example it is possible to postulate the law that equal increments in the chosen A-magnitude represent equal increments of temperature.

The mercury thermometer is an example of such a postulated law. Equal increments of the volume of mercury are deemed to represent equal increments of temperature. It is obvious that the relation of temperature to volume of mercury must be constant. If this were not true the scale would be useless. But how is it possible to tell whether the relation is constant? As Guild points out, it is impossible to determine the constancy of this relation by finding out whether various other phenomena bear a constant relation to temperature as defined by the mercury scale. That reasoning ". . . is based on an *a priori* assumption of the constancy of natural laws" (13). Guild's answer to this question is, "The point is that the constancy of the law defining our scale does not require confirmation. It is not an *assumption* which may or may not be true, it is a *postulate* forming part of the conventional framework of physical measurement. The postulated law is necessarily always true for the simple reason that it serves the purpose of defining temperature as 'the thing for which this law is true.' There is no criterion of the magnitude of a temperature (nor of any B-magnitude) other than the law by which we choose to define it. It would therefore be meaningless to ask whether the temperature to which our scale assigns the numeral n is in fact the same temperature at all times and places"(13).

The physicists seem to wish to restrict the term measurement to those magnitudes that may be measured fundamentally, i.e., A-magnitudes. For example Guild says (13), "The fact that there is no operation of addition applicable to temperature *qua* temperature, prevents it from being measurable in the true sense of the term." But it should be noted well that he also says (13), "When once we have defined some such scale of temperature, temperature becomes 'measurable' in the broad sense in which this word is generally used; and the laws relating other physical variables with *temperature as so defined* become open to empirical investigation."

The author of this study thinks that the words "open to empirical investigation" might also have been put into italics.

Before closing this section it might be well to give Campbell's definition of zero magnitudes although any discussion will be postponed to Section C.

The system B has the magnitude O, when $A = A'$, if

$$A + B = A' \qquad \text{XII}$$

SECTION C. THE OPERATIONS NECESSARY FOR FULFILLING THE PHYSICISTS' REQUIREMENTS FOR MEASUREMENT

In the section above no stress was laid on the necessary operations for meeting the criteria for measurement that were discussed. The discussion of the criteria and the discussion of the operations have been separated only for convenience. Actually the criteria and the operations by which the criteria are satisfied are inseparable. It would be possible to *describe* the criteria that must be met, it would also be possible to *describe* the

operations that ought to be performed on a group of systems in order to measure them, and still it might be impossible to measure the systems in respect of the given magnitude, *because the operations could not be carried out in practice.* The criteria are not theoretical, they are practical. The relations stated in them must be shown to hold empirically.

In glancing back at the criteria it will be noted that the following symbols were used: $>$, $<$, $\not>$, $\not<$, $=$, $\neq$, ( )[10] and $+$.

Of these symbols, ( ) and $+$ may be described as operations and the rest as relations. That is $>$, $<$, $\not>$, $\not<$, $=$ and $\neq$ state that the relation "greater than," "less than" or "equals" has been found to exist or not to exist between any two systems. While it is true that these symbols do not represent operations, they imply that certain operations have been performed on the systems so that this relation could be determined.

For example, in order to determine whether the relation $>$ existed between two systems with respect to any magnitude it would be necessary to:

1) State the operations by which $>$ is to be defined.

2) Actually perform these operations.

3) Judge whether the operational criterion has been met.

If the magnitude were weight these three steps might be applied as follows:

1) Heavier than is defined by placing two objects on a balance, one on each pan. If one of the pans sinks, the weight in that pan will be deemed to be the greater; or, in other words, the system in that pan will be greater than the system in the other pan in respect of the property weight.

2) It is now necessary to find a balance, obtain a group of systems, place pairs of them on the balance; or, in other words, actually carry out the operations used to define the magnitude.

3) There must, of course, be some way of determining whether the pan sinks or does not sink. There may be several ways of determining this fact but all of them will ultimately rest upon a judgment made by the experimenter. *The usual judgments are those of "difference" (which can be either "greater than" or "less than," or "no difference.") It should be noted again that the judgment of "no difference" is not the same as the judgment of equality. The judgment of no difference is implied in III. But in order to establish equality, it must be demonstrated that IV and V also hold.*

The question may now be asked, why does one choose one operation rather than another? How, for example, does one know that the weights should be put in opposite pans of the balance? Why not place one weight in the pan of the balance and the other on the floor? Could not this operation define the relation "greater than" for the magnitude weight?

The answer is, simply, "Try it." Suppose "greater than" is defined in this fashion. It would soon be apparent that the relation established between the systems by these operations is not asymmetrical, though it is transitive. In other words, if the operations are actually tried, it will soon become evident that it is impossible to obtain both an asymmetrical and a transitive relation. In short, weight as defined by this operation is not a magnitude. The "correct" operations are those by means of which the necessary relations may be experimentally demonstrated. The "correct" operations may be found if the experimenter is ingenious and patient. Furthermore some magnitudes now thought to be

[10] The parentheses refer to a single system $=$ to the sum of the systems included within them.

fundamentally measurable may turn out not measurable; others now not measurable will turn out to be measurable, when some ingenious experimenter discovers the "correct" operations.

>, <, etc., refer to relations that are based upon operations for their establishment but + refers directly to an operation. ( ) also refers to an operation, or, perhaps better, to the result of several operations.

( ) is used in the sense that (A + B) = that single system that is equal to the combined systems A and B. It is clear that ( ) refers to the result of several operations one of which is addition.

The operation for "addition"[11] is the greatest single stumbling block to measurement in physics and psychology. In fact the physicists claim that measurement of sensation must almost always fail because the psychologist can hardly ever find a proper operation for +. Smith (36), Cohen and Nagel (9) and Johnson (28) have stressed the fact that this is not only true for sensation but also for mental testing, the measurement of attitudes, etc.

It will be well worth while to examine the objections raised by the physicists, as it will shed a good deal of light on the criteria for additivity. Comment on the application of these criteria to the measurement of sensation and to the field of mental testing will be withheld until a later section.

The most important criterion that the psychologists fail to meet is that of "physical juxtaposition." An example of physical juxtaposition would be placing the systems end to end in measuring length; placing the weights on the same pan of the balance in measuring weight; connecting resistances in series in measuring electrical resistance, etc. Campbell (14, 7) would admit that the simultaneous presentation of two auditory stimuli to different ears would satisfy this criterion of physical juxtaposition. So too brilliance might be "added" by allowing the light from two sources to fall on the same surface. Having met this first criterion of physical juxtaposition it would then be necessary to show that VIII, IX, X and XI held. As Johnson (28) points out, they certainly do not hold for brilliance in all cases.

It seems that it was this criterion of "physical juxtaposition" that was the stumbling block to any agreement between the physicists and psychologists on the Committee of the British Association for the Advancement of Science. The Committee reported that agreement seemed unattainable on the question of whether it was possible to make quantitative estimates of sensory events because, to quote Bartlett (8), "If all measurement must conform to the Laws of Measurement enunciated by Dr. Campbell, and, in particular, if the second law can only be satisfied by the physical juxtaposition of equal entities, then sensation-intensity cannot be measured."

Where has this new requirement for measurement come from? It will be remembered that it is not discussed with the logical requirements for measurement. Although this requirement might be called an "operational requirement," there certainly must be some logical basis for its inclusion as one of the criteria that the operation of addition must meet.

It is impossible to find a clear concise statement of what is meant by "addition" in any of the publications of the

[11] In the following discussion it must be remembered that addition may refer to a logical concept or to an actual set of physical operations. We are interested in addition in this latter sense. When the word is used in this sense it will be set in quotes.

physicists that have been mentioned. It is impossible to find a single italicized sentence beginning "Addition is . . ." However it is possible to combine several statements made by Campbell and obtain a good idea of what he considers addition to be. But first it can be said that it is obvious that "addition" is no single operation that may be applied to any and all systems. The operation of adding lengths will not apply to weights.

Following is a selection of relevant phrases from Campbell:

Lord

1) . . . the systems to be measured must be capable of a certain kind of combination, which will be termed *addition* . . . (6).

2) . . . A + B means the composite system formed by combining the systems A and B in a particular way; thus, if A and B are rigid bodies, A + B may mean the body obtained by connecting them rigidly (6).

3) The conditions that any proposed form of combination must satisfy in order that it shall be addition, and shall be suitable for the fundamental measurement of any magnitude, can then be expressed in a series of propositions involving the symbols +, ( ) and >, <, = characteristics of the magnitude (6).

4) The following are the chief of these conditions.[12] They are similar to the arithmetical "laws" of commutation and distribution in addition . . . (Campbell here sets out the equivalent of VIII, IX, X, XI [6].)

5) So much for the properties of Numbers in virtue of which addition and subtraction are applicable to them. What is the similarity between these properties and the properties of bodies in respect of weight which enable us to apply to weight the process of addition? The similarity is between the relation denoted by the sign of addition and a relation which can be established experimentally between bodies in virtue of the fact that they have weight; the propositions which are true of one relation are true of the other. . . . Then corresponding to the arithmetical proposition that, if a = b and b = c, then a = c, we shall state that, if a certain body A balances another body B and if B balances another body C, then A must balance C; corresponding to the distributive law, a + (b + c) = (a + b) + c, we shall state that if P is a body which balances B and C on the same pan and Q a body which balances A and B on the same pan, then A and P on the same pan must balance C and Q on the same (plan, *sic*) pan; and so on for the other laws.

Now these statements concern experimental facts; they assert that, in certain circumstances, we shall observe something. The statements may be true or false; and, as with all statements of experimental fact, experiment only can determine whether they are true or false. If they are true there will be a certain similarity between the arithmetical process of addition and the arithmetical relation of equality on the one hand and the physical process of addition and the physical relation of equality on the other; if they are false, there will not be this similarity (5).

6) The only properties measurable *directly* by means of this rule are those (roughly termed quantities) which are *additive*—that is to say, which are such that, given two things A and B having the property, it is possible to produce by a precisely determined operation (combination) a thing C which is greater in respect of the property than either A or B . . . (6).

It is possible to gather quite clearly from the above quotations that Campbell believes that no one operation is necessarily "addition" but only those operations that experimentally fulfill the criteria VIII, IX, X and XI. In other words, an operation is "additive" if it fulfills the criteria for additivity, just as a satisfactory operation for producing the relation "greater than" is one that satisfies I (II) and II (II).

But still there is no answer to the question "why the criterion of physical juxtaposition?" Since there is no mention of it in Campbell's treatment of the logical requirements, the suspicion arises that this criterion rose after the physicists had successfully measured a number of characteristics of their

[12] I.e., the conditions mentioned in the sentence above.

systems. If the physicists found that the operations that fulfilled the necessary requirements for additivity seemed to involve "physical juxtaposition," it is possible to imagine that they *induced* that "physical juxtaposition" was a necessary requirement. It should be pointed out however that it is possible for this to be true for the systems with which the physicists deal without being a general law of measurement.

Campbell says that "Addition is a process which is peculiarly characteristic of Numbers"(5). Objective number can be scaled fundamentally (6). The result is the numeral series 1, 2, 3, etc., associated with the objective numbers 1, 2, 3, etc. Numerals and numbers have become so inseparable in our thinking that statements like this are very confusing. Perhaps it might be better to say that numerals 1, 2, 3, have been assigned to the objective numbers in the following groups of objects ●, ● ●, ● ● ●; thus, the number of this many objects, ●, has the numeral 1 assigned to it; this many objects, ● ●, has the numeral 2 assigned to it; this many objects, ● ● ●, has the numeral 3 assigned to it, etc.

The defining operations for establishing the relation "greater than" for objective numbers are: 1) select two groups; 2) pair off object for object; 3) if one group is exhausted before the other, the group that has not been exhausted is termed "greater" with respect to the magnitude "objective number." If both groups are exhausted simultaneously they are called "equal," etc. It can readily be seen that these operations establish an ordinal scale which fulfills all the necessary criteria.

But how are the groups added with respect to the magnitude "objective number" when an extensive scale is constructed? Suppose we had a group with this number of objects, ●, called A, and have found a group with an equal number of objects, ●, called A′, and we wish to combine these two groups. How do we "add" them, ● + ●? It is of course possible to place them in physical juxtaposition. So suppose that we place them on the table so that they are touching and they look like this, ●●. This added group, A + A′ is now called B. We can go ahead and find a group that is equal to B, which we call B′, etc. Now it is absurdly obvious that by using this definition of "addition" and by using the operations outlined above for obtaining the relations of >, <, etc., an extensive scale can be constructed for the magnitude "objective number." All of the criteria of additivity can be met.

It is equally absurdly obvious that we do not have to place the objects in physical juxtaposition to reach the same result. We can show that ● + ● = ●●, that ● + ● = ● + ●, etc. In fact if some of the objects happen to be in New York and others in London the scale could still be constructed and the necessary criteria could be met. It is not physical juxtaposition that solves our problem, it is the simple fact that the operations that have been adopted allow us to meet the necessary criteria.

It may immediately be argued: "It is unfair to think of physical juxtaposition in such a literal fashion. Does not Campbell say that it is possible to place one weight *in* the pan of the balance and hang the other weight underneath the same pan? It is not literal physical juxtaposition that is demanded but it is an operation that allows the *combined effect* of the magnitude to be exerted in one direction. In the case of objective number it is true that the scale may be constructed if the objects or systems are

not literally placed side by side but it *is* necessary for the two systems to be taken as a group, in other words, that the two groups B and B′ are paired off against the group C as if they were a unitary or single group."

Suppose this is so, what are the criteria for combining the systems? The above argument has not answered the problem but has restated it. In measuring length, why cannot one system be placed on two uprights and the other be hung beneath it? This would certainly be physical juxtaposition and would equally certainly meet none of the necessary criteria.

The answer can only be that the operation of placing of one weight in the pan and hanging the other underneath the pan meets the necessary logical criteria while the operation of placing one length between two uprights and hanging the other underneath it does not meet the necessary criteria. When $>$ is defined by the usual set of operations for constructing an ordinal scale of length, this operation for $+$ would not meet the criteria $A + B > A'$.

The suspicion seems to be partially confirmed that the physicists have induced this extra, operational, criterion of physical juxtaposition because it has commonly occurred in the operations they have found it necessary to use in order to meet the criteria for additivity.

It would indeed be curious if physical juxtaposition was found to be necessary for "addition" in the measurement of almost all magnitudes *except* objective number.

It has been found in discussing this point that it was extremely difficult to find examples that do not appear facetious or absurd. In the example quoted above for the measurement of length it was asked why one object could not be placed on two uprights and the other hung beneath it. This procedure seems absurd. It seems like a logical contradiction. It is obvious that if lengths are to be "added" they should be placed end to end. But is it so very obvious? The author feels that it is obvious only because it is such a well known, common, everyday experience. If one takes a less familiar example it does not appear so absurd. If the question were asked "How should I 'add' electrical resistance? Should I connect the resistance in series or in parallel?" the incorrect answer does not appear quite so unreasonable.

If one then applied the same principle of physical juxtaposition to the measurement of inductance and juxtaposed the coils and connected them in series-aiding, or series-opposing, the result would not be so happy. The result of these operations for addition would yield $L_a = L_1 + L_2 + 2M$[13] for the series aiding case and $L_0 = L_1 + L_2 - 2M$ for the series-opposing case. The strict interpretation of physical juxtaposition breaks down in this case. Yet we know that by other operations inductance may be measured fundamentally.

It has been said that $A + B > A'$ (VIII) must be demonstrated. Again it might be argued that it is obvious that the magnitude *length,* defined by the appropriate operations for obtaining the relations $>$, $<$ and $=$, cannot be "added" by hanging one object under the other. VIII can obviously not be demonstrated. But it is this very obviousness that confounds the thinking concerning the underlying logic of the problem. Really the above method of "addi-

[13] L_a stands for total inductance series-aiding and L_o for the total inductance series-opposing. L_1 stands for inductance of coil 2. L_2 stands for inductance of coil 2 and 2M is the mutual inductance where $L_1 = L_2$.

tion" should not be obviously false until it has been experimentally demonstrated to be false. It is also obvious to the electrician that electrical resistance, to be "added," must be placed in series, though it may be doubted whether this insight is inherited!

This does not mean that the experimenter may not save himself time, energy and embarrassment if he uses his intelligence and his past experience in selecting an operation for "addition" that has some hope for success. But it does mean that the final, in fact the only test, is an empirical one. Nothing but experiment can determine what operation shall be "additive" for any given magnitude.

It is now possible to attempt a definition of "addition." *Given a magnitude previously defined by appropriate operations for the establishment of the relations* $>$, $<$ *and* $=$ *between the systems with respect to this magnitude: "addition" is that operation, or series of operations, performed upon the systems in such a fashion that it is possible to meet the logical criteria for additivity (VIII, IX, X, XI).*

It should be stressed that the same operations for demonstrating $>$, $<$ and $=$ in defining the magnitude (constructing the ordinal scale) must be used in demonstrating the criteria for additivity. For example, if $=$ is defined in one way for the construction of the ordinal scale, the same definition must be used in criteria VIII, IX, X, XI, for additivity, etc. Furthermore, as Guild points out (13) the same definitions must be applicable for all parts of the same scale. There cannot be one definition for the smaller values of the magnitude and another for the larger.

SECTION D. STEVENS' POSITION[14]

Stevens defines three terms, numerals, numerousness and numerosity. He means by *numeral* a sign made on a piece of paper or an arbitrary symbol, which is the definition that has been adopted in this study. By *numerousness* is meant the property that "we discriminate when we regard a collection of objects." In this sense *numerousness might be called "subjective number." "Numerosity* is a property defined by certain operations performed on groups of objects." By numerosity he means what has been called *objective number* in the preceding discussion. He uses the term *numerosity* instead of *number* because he wishes to emphasize the difference between *numeral* and *number.*

He begins by saying that it is possible to establish a rank order of groups of objects (beans, for example) in respect of numerousness[15] simply by looking at the piles of beans and judging which of the piles is largest, etc.

"We know from experience, however, that greater reliability can be had if we rank-order the groups by pairing successively one bean from each group until one group is exhausted. Then if any beans remain in the other group, that group is said to have the greater *numerosity.*"[16] "If the pairing exhausts both groups simultaneously, their numerosity is equal. . . . Now if we designate each of these piles by a separate sign, and if we decide to use the same sign to designate all groups showing the same numerosity, we find ourselves in possession of a series of numerals. The 'spatial' (topological) order in which we write the numerals depends on the degree of numerosity designated by each numeral. . . . When we regard the numeral series as originating in this fashion, we are

[14] This summary of Stevens' "position" is gathered from (38), (40), (42).

[15] Subjective number.

[16] Numerosity is here synonymous to what has in this paper been called objective number.

not astonished to find that it exhibits certain important properties. The relations obtaining among groups, considered from the point of view of numerosity are reflected in the relations obtaining among numerals—but with an important difference: degree of numerosity among groups corresponds to 'spatial' relations in the numeral series. Likewise, the numerosity achieved by combining the groups (addition) corresponds to the numeral arrived at by stepping off two successive 'distances' along the numeral series, and since the order for combining the groups and for stepping off the 'distances' is immaterial, we can demonstrate the associative, commutative, and distributive laws both for groups of objects and for our series of numerals. Furthermore, to the extent that we can show transitivity, asymmetry, etc., among relations of numerosity, we can also show them among the topological 'spatial' relations in the numeral series. . . ." (40)

Stevens then mentions three kinds of scales for numerousness.

1) An *ordinal* scale could be set up by arranging the groups of beans in an order so that every group was either greater or less or equal to every other group.

2) An *intensive* scale could be devised by prescribing a semantical rule for determining the assignment of adjacent numerals; for example, adjacent numerals could be assigned to groups that showed a just noticeable difference in numerousness.

3) An *extensive* or additive scale could be constructed by determining when one group appeared one half as numerous as another. "If this judgment could be made, we could then assign to the smaller group the numeral lying in the numeral series midway from the beginning of the series and the numeral assigned to the larger group. If we assigned numerals according to this procedure, the relations among groups exhibiting the property numerousness would be reflected in the spatial relations of numerals within the numeral series"(40).

Stevens (38), in 1936, discussing the problem of scaling in psychology, claimed that the purpose of scaling was to facilitate the description of natural phenomena by means of functional relations expressed, if possible, by the conventional mathematic symbols. In order to accomplish this it is desirable to assign numbers[17] which not only denote the order of the systems but also the relative magnitude of the phenomena. "When this is done, the scale numbers can be manipulated in accordance with arithmetical laws in order to determine additional relationships such as the sum of two magnitudes, . . . etc. However, the outcome of the purely formal (mathematical) manipulation of the scale numbers has no significance unless the manipulations and their results can be identified with some concrete operations. First the scale numbers[17] should be applied to the attribute of sensation in such a way as to make the scale one of true numerical magnitude, which means simply that if the numbers are manipulated according to the rules of arithmetic, the result (and the manipulations) correspond to a set of physical operations. Secondly, although at the outset we could conceivably choose any one of several sets of operations as defining the scale, that set will ultimately prove to be most satisfactory for a subjective scale when it leads to scale numbers bearing a reasonable relationship to the experience of the observer." He continues by saying that a scale would be satisfactory if the magnitude[18] of a particular discriminable characteristic to which the

[17] What we have here called numerals.
[18] I.e., subjective magnitude.

numeral 10 had been assigned was half as great subjectively as that to which the numeral 20 was given and twice as great as that to which the numeral 5 was given. "With such a scale the operation of addition consists of changing the stimulus until the observer gives a particular response which indicates that a given relation of magnitude has been achieved." He says again, "A scale, then, which would enable us to designate the *numerical* as well as the *intensive* magnitude of an attribute of sensation can be constructed according to the criterion that, having assigned a particular number N to a given magnitude, the number N/2 shall be assigned to the magnitude which appears half as great to the experiencing individual."

The actual method of assigning numerals in order to construct a scale of this sort is somewhat more confusing than the above statement would indicate. As this method of scaling will be referred to fairly frequently in the subsequent discussion, it will be outlined here in some detail. Other accounts may be found in 39, 40, 42.

In a typical case the subject is presented with a standard stimulus and adjusts a variable stimulus until it appears to be half the subjective magnitude of the standard. These 1/2 judgments are obtained for a number of standards which have been selected to cover a wide range of physical magnitudes. It is then possible to plot the stimuli judged 1/2 against the standard stimuli. The units used in this plot are physical measures of the stimuli, such as frequency, centimeters, etc. A curve is then fitted to the obtained points. This plot is referred to hereafter as the half judgment function. Figure 1 shows such a plot for a group of imaginary data. In this figure the stimuli judged 1/2 are

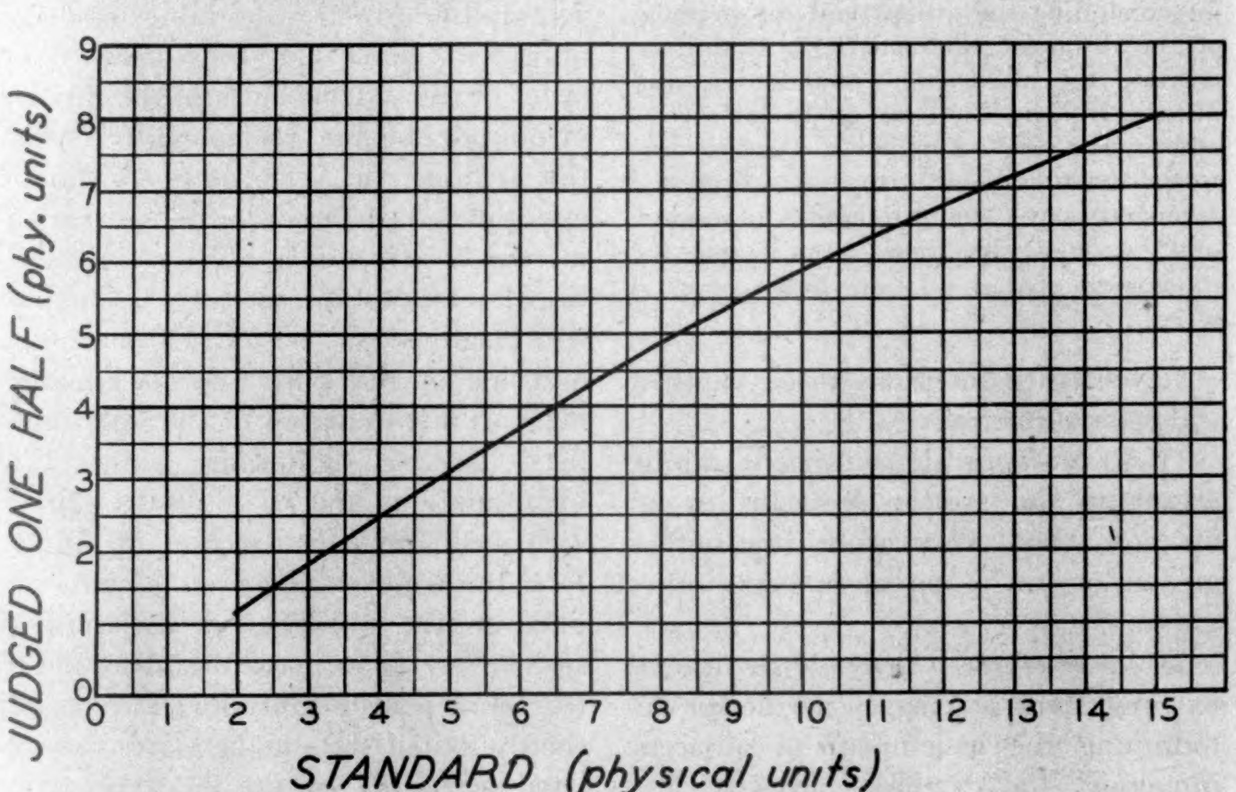

Fig. 1. Hypothetical half judgment function.

plotted, in physical units, against their respective standards, also measured in physical units.

The magnitude function or scale is constructed from this plot in the following way:

1) A numeral is arbitrarily chosen and assigned to some subjective magnitude associated with some arbitrarily selected stimulus. In this case the numeral 1 has been arbitrarily chosen and has been arbitrarily assigned to represent the magnitude of the discriminable characteristic associated with the stimulus physical magnitude of 1. This fixes the first point of the magnitude function (Fig. 2), indicated by a cross.

2) Returning to Figure 1 it is seen that a stimulus of 1 physical unit was judged by the subject to be 1/2 the subjective magnitude of a stimulus of 1.75 physical units. *In other words the magnitude of the discriminable characteristic associated with a stimulus of 1 physical unit is judged to be 1/2 as great as the magnitude of the discriminable characteristic associated with a stimulus of 1.75 physical units.*

Since the magnitude of the discriminable characteristic associated with a stimulus of 1 physical unit is judged to be 1/2 the magnitude of that associated with one of 1.75, and since the magnitude of the discriminable characteristic associated with a stimulus of 1 has been assigned the numeral 1, then the numeral 2 (2 subjective units) is assigned to the magnitude of the discriminable characteristic associated with a stimulus of 1.75 physical units. This point is then plotted (Point A, Fig. 2).

3) Returning to Figure 1, it is necessary to find of what discriminable characteristic the discriminable characteristic associated with the stimulus of 1.75 physical units is judged to be 1/2. It is seen that the discriminable characteristic associated with the stimulus of 1.75 physical units is judged to be 1/2 of the subjective magnitude of the discriminable characteristic associated with a stimulus of 2.85 physical units.

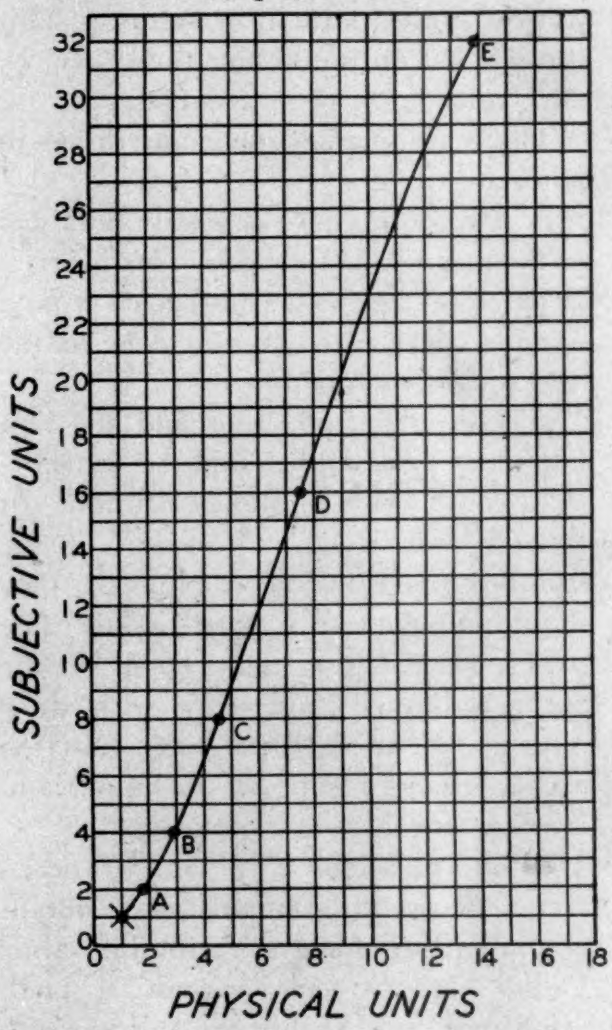

FIG. 2. The magnitude function constructed from the hypothetical half judgment function presented in Figure 1 (arithmetic coordinates).

Since the magnitude of the discriminable characteristic associated with a stimulus of 1.75 physical units has been judged to be 1/2 the magnitude of one of 2.85 physical units, the numeral assigned to the stimulus of 2.85 units must be twice the numeral assigned to represent the magnitude of the discriminable characteristic associated with a stimulus of 1.75 physical units. Two subjective units (the numeral 2) have

been assigned to the magnitude of the discriminable aspect associated with a stimulus of 1.75 physical units so 4 subjective units are assigned to represent the magnitude of the discriminable aspect associated with a stimulus of 2.85 physical units, point B in Figure 2. This process is continued for points C, D, etc., until the scale is complete.

There are several important things to note:

1) A smooth curve is always obtained in the magnitude function. This is necessarily so because it is constructed from a smooth curve fitted to the data in the half judgment function.

2) The magnitude plot and the half judgment function should be extrapolated only with extreme caution. In fact, if only a limited section of the stimulus range has been explored, they should not be extrapolated at all. The reason for this statement, based on the author's limited experience with these plots, is that the half judgment function usually changes slope as the upper and lower thresholds are approached.

3) Interpolation is obviously necessary, as it would be impossible to obtain 1/2 judgments for every stimulus value. Values closer together than 1 j.n.d. would not add anything to the accuracy of the graph.

There can be no rule for determining the number of stimuli used. Common sense can be the only judge. It should be clear, however, that the stimuli should be closer together at critical points on the curve. Critical points would be those where the curve changes slope rapidly or near the point of break in a discontinuous function.

4) In the example of scale construction used above, the arbitrary starting point was the lowest stimulus value. The magnitude plot could have been obtained by arbitrarily starting at the highest stimulus value or even by starting with a stimulus value that is in the middle of the stimulus range. This latter procedure was adopted by Stevens and Volkmann (42).

If the arbitrary starting point is not at the bottom of the stimulus range a slightly different procedure must be adopted in constructing the magnitude plot.

Suppose a stimulus of 4.60 physical units had been chosen for the arbitrary starting point, then:

1) A numeral is arbitrarily chosen, as before, and assigned to the magnitude of the discriminable characteristic associated with a stimulus of 4.6 physical units. Assume that the numeral 8 has been assigned to the stimulus of 4.6.

All the numerals for stimuli above 4.6 physical units are obtained as outlined above. Numerals for stimuli that are lower than 4.6 physical units are obtained by asking the question, "What stimulus was judged to be 1/2 of 4.6?" To answer this question it is necessary to go to Figure 1 and go out the abscissa to 4.6 and up the ordinate to the value of the stimulus judged 1/2 which in this case is 2.85.

Since the discriminable characteristic associated with a stimulus of 4.60 physical units was assigned 8 subjective units, then the magnitude of the discriminable characteristic associated with 2.85 should be assigned 4 subjective units.

The question is then asked, "What stimulus was judged to be 1/2 of the stimulus of 2.85 subjective units?" and the scale is continued in this fashion until complete.

In this example, where the numeral 8 is assigned to the stimulus 4.6, the resulting scale would be identical with the scale constructed previously with the

arbitrary starting point at the bottom of the stimulus range, i.e., with the numeral 1 assigned to discriminable characteristic associated with the stimulus of 1 physical unit. The reason for this is that the numeral 8 is the numeral that would have been assigned to the stimulus 4.6 *if* the scale had been started from the bottom (stimulus 1) with an arbitrarily chosen numeral of 1. If any other numeral than 8 is assigned to the stimulus 4.60 the two scales will not be identical. *However the relations between the stimuli as expressed by the scale numerals will be the same no matter what the arbitrary starting point or what numeral is assigned to it.*

There is another set of operations used for constructing a magnitude function in psychology when the data have been obtained by the method of equal appearing intervals. The method of equal appearing intervals itself is too well known to require description here. But the technique of construction of the function after the data have been obtained differs from the method used for fractionation data. Although mention will be made of the method of equal appearing intervals in the present paper, no use will be made of the technique for the construction of the magnitude function with data obtained from the method. The reader is referred to Stevens and Volkman (42) where a good description of the method may be found.

Stevens and Volkmann (42) argue that the results obtained by this method should check the results obtained by the method of fractionation. This of course does not mean that the values (numerals or units of subjective magnitude) assigned to any stimulus by the scaling technique employed after the use of the method of fractionation should be identical with the values assigned to the same stimulus after the use of the method of equal appearing intervals. It does mean that the relations between stimuli as expressed by one set of numerals should be the same as the relations between stimuli as expressed by the other set of numerals.

SECTION E. CRITICISMS OF PSYCHOLOGICAL MEASUREMENT BY THE PHYSICISTS

There are four basic methods by which measurement has been attempted in psychology, the method involving the integration of j.n.d.'s, the method of equal appearing intervals, the method of fractionation and the various techniques based on the normal probability curve.

The objections and criticisms of the physicists will be briefly outlined for each of these methods in turn:[19]

1) The Integration of j.n.d.'s

a) The scale constructed from the integration of j.n.d.'s can not measure an A-magnitude, as the defining relations involve the measurement of another magnitude, namely stimulus intensity. The defining relations for an A-magnitude must be independent of all other quantitative relations for other magnitudes. Furthermore the only way in which the scale could measure a B-magnitude is to "define S by a postulated relation to I." The scale has some of the properties of both A and B magnitudes but has neither the necessary nor the sufficient properties of either (13).

b) Fechner assumed that all j.n.d.'s are equal. What are the actual specified relations between j.n.d.'s? $\triangle S_1$ is the sensation increment associated with a j.n.d. at stimulus intensity I_1 and $\triangle S_2$ is the sensation increment associated with

[19] This summary of the physicists' criticism has been taken chiefly from (7), (13), (14).

the j.n.d. at intensity I_2. This statement is the only specified relation between $\triangle S_1$ and $\triangle S_2$. This relation is not symmetrical, as it ceases to be true when $\triangle S_1$ and $\triangle S_2$ are interchanged. But to establish equality the relation must be shown to be symmetrical. "This one consideration alone renders superfluous all the semimetaphysical arguments which have centered round the question whether or not equal, in the sense of equally noticeable, necessarily means 'really' equal. A symmetrical transitive relation is essential as a practical criterion of equality in measurement" (13).

c) The criterion of equality is not applicable throughout the scale. It is meaningless to say that

$$S = \Delta S_1 + \Delta S_2 \ldots\ldots\ldots\ldots\ldots\ldots \Delta S_n$$

because = in this case has a meaning different from that used for equality of different $\triangle$ S's (13).

There seems to be no reasonable defense against these arguments.

2) *The Method of Equal Appearing Intervals*

The chief arguments presented by the physicists against the method of equal appearing intervals are:

a) The magnitude obtained by the method of equal sense distances is not the same magnitude as *sensation intensity*. The operations for finding equal sense distances involve 3 or more stimuli of different apparent intensities while those for finding equality of sensation intensity involve two apparently equal stimuli. The operations for obtaining equality are different in the two cases, therefore the magnitudes are also different (13).

b) Nor, they argue, can it be stated in rebuttal that the above objection applies with equal validity to any magnitude. While it is true that it is impossible to obtain equal differences of length without at least 3 objects of unequal length, the analogy is not valid. As Guild says, "Difference of lengths as something expressible on a quantitative scale derives its significance from the association of number and length established by the practical criteria of equality and addition which define length as a magnitude. It merely means the length which must be added to the smaller of two lengths in order to make a new length equal to the larger of the original pair. We cannot define a process of subtraction independently of a process of addition. We cannot construct a scale of length from units of difference-of-length defined by operations other than those involved in defining equality and addition for length. Similarly we cannot give any quantitative significance to difference-of-sensation-intensity unless we already have practical criteria both of equality and addition for sensation intensity; for all that difference-of-sensation-intensity means, if it means anything, is the sensation intensity which, when added to the smaller of two given sensation intensities, will produce a new intensity equal to the larger" (13).

c) The proposed criterion of equality is inadequate, as a symmetrical, transitive relation cannot be demonstrated.

This last objection seems to be decidedly ill-founded. Given the four stimuli A, B, C, D, marking off the three sense distances, AB, BC, and CD,

A B C D

and given the fact that BC and CD have been judged equal, it can be shown, obviously, that BC $\not>$ CD and BC $\not<$ CD (III); also if BC is judged to be $>$ AB, it is more than likely that CD will be judged $>$ AB (IV). Given the case below

where BC and CD had also been judged as equal,

A B C D

it is more than likely that if BC is judged $<$ AB, CD will also be judged $<$ AB (V).

The author believes that two questions are raised by objection *b*. The first question is whether the observer can actually equate sense distances. The second question concerns the interpretation to be placed on the results if he can equate them.

What is the answer to the first question? How do we know that the observer can equate sense distances? Can this equation have any meaning apart from the operations used to define addition? It can, for the operations for obtaining the relation = are certainly not dependent upon the operations for defining addition. The criteria for = are mentioned under III, IV and V. These criteria may be shown to hold for the relation of = established by the method of equal sense distances. There can be no possible objection to the statement that the sense distances are equal, for the established equality meets all the necessary logical criteria for equality.

Why, then, is objection *b* raised at all? The difficulty seems to be with the word "difference." The physicists argue that a difference may not be defined apart from addition, i.e., the difference between A and B is that amount that must be added to B to produce A. In other words if $B + X = A$, then $A - B = X$, but they argue that this second proposition has no meaning apart from the first proposition. Since additivity has not been demonstrated, the concept of "difference" has no meaning in the equal sense distance experiment.

It seems to the author that this criticism is due to a certain amount of verbal confusion. It is true that the *word* difference is sometimes used to describe the sense distances equated by the observer, but this does not imply that the observer need perform two mental subtractions and additions in making his judgment. It is possible to call the interval whatever one wishes but this does not change the fact that the observer has the straight forward task of equating the intervals,

O A

O B

O C

AB and BC, so that AB = BC.[20] He is perfectly capable of making this judgment with some consistency and the resulting equation satisfies all the criteria for equality.

But even if it is granted, and the author believes it must be, that the observer can make this judgment of equality in such a way that all of the criteria for equality are satisfied, all of the physicists' objections are not answered.

The psychologist is not interested in the simple result AB = BC. This result is but a means to an end. He desires to construct a scale, he wants to be able to say that if A = O (the absolute threshold) and the numeral 1 is assigned to B, the numeral 2 must be assigned to C.

A
O

O B

O C

[20] This is not a new viewpoint. It seems to the author that it is essentially the same as Delboeuf's *Contrastes Sensibles*.

He reasons that this is possible because C is as "far" from B as B is from A. But the physicist would argue that the numerals so assigned cannot be interpreted in the conventional fashion unless it is shown that AB + BC = A.C.

This objection seems far-fetched because B was so adjusted that AB = BC and it should follow logically that B bisects AC and as a consequence, since two halves make a whole, AB + BC must equal AC. But the criteria for measurement are not only logical, they art practical. The statement, that AB = BC, therefore AB = 1/2 AC, and therefore AB + BC = AC is a sound logical deduction but it may not correspond to any demonstrable relation between the stimuli in question. Although it is logical that the two halves make up the whole, it may not be experimentally demonstrable in a given situation. For example, suppose the introduction of the stimulus B increased the subjective magnitude of the distance AC. If the numeral 1 was assigned to B and the numeral 2 was assigned to C, the magnitude of C presented alone would be misrepresented by the numeral 2 assigned to it, e.g., suppose the observer is asked to adjust a stimulus B so that AB = BC when the stimulus A has O objective magnitude,

A
O

C
O ı

The line C has associated with its physical length a subjective length which may be called X.

Now suppose the observer begins to draw the line B which has a subjective length Y associated with it.

A
O

B
O ı

C
O ı

The observer is to make the equation OY = YX. The subjective magnitude of C may be increased because of a contrast effect. This new subjective magnitude of C may be called X + e.

Now when the observer adjusts B so that OY = YX he will be actually adjusting B so that OY = Y (X + e). If the numeral 1 is now assigned to Y, the numeral 2 cannot be assigned to the subjective magnitude of C, i.e., to X. It must be assigned to X + e.

It can not be *assumed* that subjectively AB + BC = AC simply because AB = BC. It must be *demonstrated.*

Objection *a* might be answered empirically. It was seen that properties that have the same order are considered to be magnitudes of the same kind. If, then, it can be shown that the order resulting from one series of operations for obtaining >, < and =, is the same as the order resulting from another definition of >, < and =, it can be said that the two sets of operations measure magnitudes of the same kind. It can be confidently predicted that, for most subjective magnitudes, the results of an empirical investigation would show objection *a* to be ill-founded.

3) Fractionation

a) The first objection to the method of fractionation is similar to objection *b* raised against the method of equal appearing intervals. It is not possible to define ½ without reference to addition. The observer should not be entitled to use the word ½, if the facts from which

the word derives its significance do not exist in the sphere of discourse (14).

Furthermore, granting that the observer can adjust $B = \frac{1}{2}C$, he has not demonstrated that $B + B' = C$. This can only be done if an operation for addition that meets all the necessary criteria has been found.

b) Campbell says that the monaural-binaural method of scaling loudness satisfies the criterion of additivity (7, 14). The only reason for comparing the scale obtained by fractionation with the scale obtained by this method[21] is to see whether the observer can guess ½ correctly. Even if the observer could guess ½ correctly, no one would abandon the exact physical measurement of magnitudes in favor of measurement based on guesses.

As in objection *b* raised against the method of equal appearing intervals, there are two points at issue in objection *a*.

The first is, can there be any meaning to the operation of halving unless additivity has been shown? The second question is, assuming that the ½ judgment has a meaning apart from addition, do the assigned numerals give an extensive scale?

To answer the second question first, it would seem that it is necessary to remember that the assigned numerals can not express any relation that has not been demonstrated empirically. Stevens says, "If we assigned numerals according to this procedure,[22] the relations among group exhibiting numerousness[23] would be reflected in the 'spatial' relations of numerals within the numeral series." This is true. *But the relation* $B + B' = C$ *is not one of the relations demonstrated by the fractionation experiment. By the very nature of the experiment the only relation claimed to be demonstrated is* $B = 1/2C$.

Certainly no objection can be raised to Stevens' theoretical position concerning the origin of the numeral series and the criteria that must be met to be able to obtain an additive scale. Nor can there be any objection to the statement quoted immediately above. The only question at issue is whether the scale constructed by the method used by him can be accepted as additive when the additive nature of the magnitude has not been demonstrated. The magnitude may be additive, but it has not been shown to be so.

In answer to the first question, it is necessary to note that there are certain similarities and differences between the method of equal appearing intervals and the method of fractionation. In order to examine these it might be helpful to diagram the method of equal appearing intervals in a way different from that used above when stimulus A = O.

O ———————————— A

O —————————————— B

O ———————————————— C

This diagram gives a somewhat clearer picture of the true situation. These are three stimulus intensities; one, A, has the absolute magnitude represented by the distance from zero (the absolute threshold); the second, B, has a greater absolute magnitude, represented by the distance from zero to B; and the third, C, a still greater absolute magnitude, represented by the distance from zero to C.

[21] He is discussing the comparison made in Stevens and Davis (39a).

[22] I.e., the procedure used after a fractionation experiment.

[23] I.e., subjective number.

The fractionation experiment may be diagrammed as follows:

A
O
O ———————————— B
O ——————————————————— C

The important difference to note is that in the fractionation experiment the stimulus A is zero. Aside from this difference it is immediately apparent that the two situations are strikingly similar. In fact it may be said that the stimulus situation in the fractionation experiment is a special case of the equal appearing intervals experiment.

What is stimulus A which is zero? It simply means that no stimulus is given, it being assumed that the observer has some idea of what zero magnitude is like. Stevens and Volkmann (42), in a successful attempt to improve the reliability of their results for the pitch function, introduced a stimulus of zero pitch. It might seem that the presentation of a stimulus of zero magnitude must mean the presentation of no stimulus at all. This is not necessarily true. The experimenter does not present a *stimulus* of zero physical magnitude; he presents a stimulus of such an intensity that the characteristic being scaled has zero subjective magnitude. There may be other characteristics of the stimulus that have a subjective magnitude above zero. But still it would seem that, in effect, the experimenter was presenting a characteristic that was subjectively non-existent; hence, so far as that characteristic was concerned, he might just as well have presented no stimulus at all. The observer is no better off than if he had supplied his own idea of zero. This is true. In practice, then, the zero stimulus is not really a zero stimulus but one that is just noticeably above zero. This procedure was followed by Stevens and Volkmann (42) and was found to be helpful for some of the observers.

It should be noted that the "almost zero" stimulus was not presented to the observer in the usual way. In reality it was available for him to use if he needed it, in order to clarify his idea of what zero was like. He could turn it on or off at will.

It may be argued that this procedure of allowing the observer to hear a zero characteristic that is not really zero will introduce a constant error into the results. Stevens and Volkmann recognize this possibility but argue that the effect of this error would be negligible for the high tones, and the subject was not allowed to use this "almost zero" stimulus for the fractionation of the lowest tone. Furthermore, when the almost zero stimulus is used, the resulting increase in accuracy will undoubtedly more than compensate for any small deviation caused by its introduction. Also the possibility of a constant error will be reduced progressively as the "almost zero" stimulus approaches the absolute threshold.

Stevens and Volkmann (42) point out the possibility that the introduction of a zero stimulus may turn the fractionation experiment into an equal appearing intervals experiment. They reason that logically the fractionation to one-half is equivalent to the bisection of some higher point and zero. They say, however, that in their experiment the fractionation procedure differed in two important respects from the equal appearing intervals experiment. The first difference is that the observer was not given the A stimulus but it was merely available if he wished to use it; the second difference is that the observer was instructed to use the "almost zero" tone (A stimulus) as a reference tone and not "as a limiting tone in a self imposed task

of bisection" (42). The second difference would seem to be the more important, for it seems that even if there is no A stimulus presented at all, the subject may still make an equal appearing interval experiment out of the fractionation experiment by self-instruction. It is true that what the experimenter does is important but it also is true that what the subject does is very often more important. Even if the experimenter does not allow the observer to use an A stimulus, the observer may still equate the intervals AB and BC. He can supply his own A stimulus. It must always be remembered that there are always two kinds of operations in an experiment of this sort, those that are under the direct control of the experimenter and those that are under the direct control of the observer. No matter what the experimenter does, the operations of the observer can often determine the nature of the experiment. It can, in fact, be stated that the most important differences between these two methods are those that depend on the operations not under the direct control of the experimenter.

It will be worth while to examine the difference between the two experiments in more detail.

In the equal appearing intervals experiment the observer,

1) Must disregard the absolute magnitude of the stimuli.

2) Equate the distances AB and BC.

The question as to whether the observer can halve or bisect the distance AC is the same question as whether he can halve the stimulus in the fractionation experiment.

In the fractionation experiment the observer might do one of three things:

1) He might equate the distance between AB (when A = O), and BC.

2) He might set B to one-half the absolute magnitude of C *(this is what he is told to do in the instructions).*

3) He might adjust the absolute magnitude of B so that B + B = C.

If the observer chooses the first of these possibilities he automatically performs the operations inherent in the method of equal appearing intervals.

With respect to the second possibility, can the observer find ½ the absolute magnitude of C? This is essentially the question raised in objection *a.*

The author believes that the observer can do this in only two ways. The first way is to define ½ C as that position of B when AB = BC. In other words, the observer can perform this task when he changes it into an equal appearing intervals task. The second way is to define ½ as that position of B that allows the absolute magnitude of B added to itself to equal C. In other words, B + B = C. This is the third operation mentioned above. In other words, it seems that the concept of ½ can have no meaning apart from these operations. But the physicists would argue that it has no meaning other than that which is dependent on addition.

It is, then, necessary to examine the fractionation method of obtaining half to see if the ½ so defined is independent of addition.

It is true that the observer might define half as that position of B that divides the distance AC into halves. But what can the observer actually do to obtain this relation? All he can do is to equate AB and BC. He makes judgments of "difference" and "no difference" on the AB and BC intervals. This has nothing to do with "halving." It is the operation for obtaining equality. But where then does the half concept arise? It arises because the observer knows that a point that bisects a distance logically divides that distance into two halves. But not only is a half one of those two

equal parts into which a whole may be divided, it is also one of the two equal parts which actually add up to make the whole. And the argument is right back at the beginning. The concept $\frac{1}{2}$ cannot be defined apart from addition. It is true that the observer can make the logical deduction that if the whole is divided into two equal parts the two parts must be equal to the whole. Again, however, we are up against the fact that the criteria for measurement are practical; they are not assumptions. The relations between the systems must be demonstrated facts. The statement that $B = \frac{1}{2} C$ is meaningless unless it can be demonstrated that $B + B = C$.

The only conclusion that can be drawn, then, is that the fractionation experiment becomes an equal appearing intervals experiment if the observer defines $\frac{1}{2}$ as that position of B which makes $AB = BC$. The results must be interpreted in the same way as those obtained by the equal sense distance method and they are subject to the same limitations.

Suppose the observer defines $\frac{1}{2}$ as that magnitude of B which fulfills the conditions that $B + B = C$. It must be noted that this definition can only be made with reference to the absolute subjective magnitude of the stimulus B. If the absolute magnitude of B added to itself equals C, then one can say that $B = \frac{1}{2} C$. It is impossible to define $\frac{1}{2}$ by adding the sense distances rather than the absolute magnitudes, for if $AB + BC = AC$, it cannot be deduced that B bisects the distance AC unless it has previously been shown that $AB = BC$. There is some evidence that this definition may be used. One observer has reported that he has used this definition of $\frac{1}{2}$ or, rather, it is truer to say that he used both definitions of $\frac{1}{2}$. This observer reported that he checked one definition against the other, i.e., he would equate the distances AB and BC and then to check the equation would ask himself the question, "Does the absolute magnitude of $B + B = C$?" It may be that some observers have as much difficulty defining $\frac{1}{2}$ apart from addition as the logicians think they should have.

The fact that even one observer has reported a subjective additive operation is extremely interesting. In so far as any observer uses the additive definition of $\frac{1}{2}$ the objection that $\frac{1}{2} C = B$ does not mean that $B + B = C$ loses its validity. This does not mean that the fractionation method gives us an additive scale. To obtain an additive scale it is necessary that the subjective operation of addition used by the observer meet the necessary criteria and, furthermore, it is of course necessary to be certain that the observer actually does use the additive definition of $\frac{1}{2}$ and not the equality definition. Although an additive scale has not yet been demonstrated, the way seems to be open by which the psychologist may attempt to demonstrate additivity for the magnitudes that he "measures." He is free of the false criterion of "physical juxtaposition" and may go about his business of attempting to find those operations that will meet the logical criteria for measurement.

There remains one further possibility to be discussed. That is the possibility of additivity in the equal appearing intervals experiment when A is not O. It is now obvious that the observer cannot *halve* or *bisect* the sense distances with any meaning unless he adopts an additive process. He can, of course, equate the distances but as has been seen this does not of itself give an additive scale. If the observer were to adopt an additive operation in the equal appearing intervals

experiment, he would have to restrict himself in the following manner, "If B is to bisect the distance AC, it would mean that the absolute difference AB added to the absolute magnitude of B should equal C." This is obviously an elaborate and confusing instruction. The fractionation technique offers a simpler way of obtaining the same results. Furthermore we find the word "differences" creeping into the instruction and it was seen that additivity must be shown before the concept of difference can have any meaning.

Stevens and Volkmann (42) compared the results obtained by the fractionation method and the method of equal appearing intervals and found that the two methods yielded essentially the same scale. It is, of course, not known whether the observers in the Stevens and Volkmann experiment used the equality definition of ½ or the additive definition. One suspects that they used the equality definition, as this is the easier for the observer. If they did it is not surprising that the two scales checked each other as they were constructed by the same series of operations. It is true that the two methods might yield different results if the two sets of instructions operated differentially in producing either complicated or erroneous self-instructions or constant errors depending on the presence or absence of a zero stimulus.

If the analysis of the operations involved in the equal appearing intervals experiment and the fractionation experiment is correct, the lack of reliability of fractionation to ⅓ or 1/10 would be easily explained. If the observer is given a sense distance A ——— N which he is to divide into 5 equal appearing intervals, he can perform this task with some ease. If he is asked to fractionate a magnitude to 1/5, the task is difficult and the results do not check the results obtained by the other methods. In the equal appearing interval task the observer must equate 5 intervals all of which are given, i.e., the stimuli are under his control and he can compare every interval with every other one directly. In the fractionation to 1/5 this is not possible. If the above analysis is correct, the observer must either find that stimulus which when added to itself 4 times would give N; or, he must adjust B so that the interval AB is equal to 4 other intervals between A and N, *none of which are given!* One should certainly not expect very accurate or reliable results from such a procedure.

Objection *b* raised by Campbell is not really important. If the fractionation scale is not valid because additivity has not been demonstrated, the objection is superfluous. If additivity has been demonstrated, and the scale constructed by fractionation, there is no reason why the two scales should give the same results. In fact from Campbell's discussion it might be assumed that they would not. Campbell says that properties that have the same order are the same magnitude or are magnitudes of the same kind (6). However when more than order has been demonstrated, in other words when an additive scale has been constructed, this statement may be limited. Campbell says ". . . the identity of magnitudes (or at least of such magnitudes as are susceptible to fundamental measurement) arises from similarity of addition, and this suggestion is correct; magnitudes are the same, however greatly they may differ in the relation $>$, if an operation $+$ can be found which satisfies the conditions of addition for all of them" (6).

4) Statistical Methods of Scaling

The author does not intend to go into

any great detail in discussing the objections raised against the statistical scaling methods. An excellent critique may be found in Smith (36).[24] There are, however, one or two important points that might be made.

It is possible to define such a magnitude as difficulty of mental test items in terms of the percentage of persons solving the items. Two items may be defined as equal in respect of difficulty if they are passed by the same percentage of people. An item may be defined as more difficult than another item if it is solved by a smaller percentage of people. Likewise it may be defined as less difficult if it is passed by a greater percentage of people. By the use of these definitions and by performing the necessary operations upon a group of items, the items may be arranged in an order of difficulty. The experimenter is in possession of an ordinal scale. He has not established equal units nor has he demonstrated additivity.

Psychologists have been very anxious to demonstrate that they could obtain equal units of difficulty for measuring mental test items. They thought that the demonstration of equality of units allowed them to add and subtract the numerals representing the items. This, as has been shown, is not true. Not only is it necessary to demonstrate the equality of the units, but it is also necessary to demonstrate that the magnitude is susceptible to addition.

Several methods of manipulating test results have been devised. One common method consists of translating the percentages of the attained distribution into units of the base line of the normal curve. These base line units are considered equal. It is true, of course, that the units of the base line are geometrically equal. But still it has not been demonstrated that these units that are geometrically equal correspond to equal units of difficulty. One unit distance along the base line is equal to any other unit distance. This does not mean that the distance in difficulty between the test items that these units represent is also equal. This relation is assumed. There is no experimental operation demonstrating this equality. *Experimentation may demonstrate a relation but mathematical manipulation cannot create one.*

Smith gives a very useful example, "Let us consider a normal distribution of people, with respect to height. It is true that equal units of the base line of the curve will mark off equal units of this quality. For example, if the mean height is 60 inches, and sigma 1 is 5 inches, then it is true that the individuals whose frequencies place them at sigma 1 will be 65 or 55 inches high, depending upon whether the sigma is plus or minus. It is also true that the frequencies may be manipulated to obtain units of height, just as they are used to obtain units of learning. But quantitative units of height are not established by this procedure. Rather, sigma 1 has a quantitative meaning because height has been shown in quite another context to be additive. Apart from the fact that units of length had already been ascertained on operational grounds, there would be no reason to assert that the sigma marked off a unit of height in a fundamental and additive sense. It is the fact that height has been measured *fundamentally,* and independently of its career in a normal distribution, that gives quantitative meaning to a segment of the base line" (36).

The psychologist who is attempting to measure difficulty is placed in a position

[24] See also Peatman (33).

similar to that of the physicist who wishes to measure temperature. It will be remembered that the physicist postulates a relation between equal increments of an A-magnitude and temperature. When the physicist has done this "the laws relating other physical variables with *temperature as so defined* become open to empirical investigation" (13).

The psychologist has just as much right to postulate a relation between difficulty and σ units or, for that matter, he may postulate a relation between difficulty and percentage passing. There is no more reason to postulate the relation that equal units of the base line represent equal units of difficulty, than there is to postulate the relation that equal units of percentage represent equal units of difficulty.

If the psychologist postulated a relation between percentages and difficulty, there seems to be little doubt that he would remember that this was not a demonstrated relation but a postulated one. Certainly no psychologist would contend that it has been demonstrated that the difficulty between two test items passed respectively by 80 and 90 per cent of the population is the same increment of difficulty that exists between two items passed by 50 and 60 per cent of the population. But he may certainly postulate this relation. If he does, he may then find the laws relating other variables to difficulty so defined. When the psychologist uses σ units, and has thereby gone one more step from his original data, it seems more difficult for him to realize that the relation is still a postulated one and not a demonstrated one. But, again, by postulating this relation he can provide himself with a very useful tool. He will not get into difficulty until, either in his theorizing or his practice, he forgets that the relation is not demonstrated, but postulated. When he adds these σ units he is adding equal distances along the base line of the normal curve and when he subtracts these units he is subtracting units of equal distance along the base line of the normal curve; he is not adding units of difficulty nor is he subtracting units of difficulty.

Scales so constructed may be of immense value, particularly in "applied" work. When they are used as research instruments, or when they are used as the basis for theorizing, they may well lead to faulty conclusions. That is, they may do this if the person interpreting the results forgets that he is dealing with a postulated relation. If difficulty could be measured fundamentally, it would be possible, even probable, that the relation between difficulty, so defined, and other variables would be very different from the relations between σ unit difficulty and these same variables.

A similar situation exists in the field of learning. At the present time it is considered impossible to measure learning directly. It is considered to be a B-magnitude, measured in terms of other fundamental magnitudes, such as time. It is a strong temptation to regard learning as something which exists apart from the operations used to measure it. If the psychologist adopts this attitude, he is surprised when he discovers that different measures of learning do not always give the same results. He forgets that the operations that he uses create the magnitude, and there is no reason to assume in advance that the magnitudes created by different operations should be the same.

Thus the "strength" of a reflex may be defined in terms of response latency, magnitude of response and response rate. Then, if these measures do not corre-

spond perfectly, the psychologist may be puzzled because he believes there ought to be perfect correspondence, as all measure the same thing, viz. "reflex strength." The fallacy is obvious from the preceding discussion. There may be no unitary magnitude "reflex strength" defined in terms of response latency, magnitude of response and response rate but simply three independent B-magnitudes, each one of which varies in its own characteristic fashion when it is the dependent variable in a given experiment.

SECTION F. CRITICISMS OF PSYCHOLOGICAL MEASUREMENT BY PSYCHOLOGISTS

The plan adopted for dealing with the physicists' objections will be used in this section. First the theoretical position taken by some psychologists will be outlined and then some of the specific criticisms against the experimental methods will be discussed. As was obvious in the previous section, this division is arbitrary. In reality, theory cannot be divorced from practice.

No attempt will be made to present all of the criticisms of all psychologists. Obviously this would be an enormous and useless task. In fact, only a selected group of criticisms will be offered. These will be limited to criticisms that are directly pertinent to the points raised by the logicians. Much of the writing of the psychologists has been repetitive, so that only representative arguments will be brought forward.

McGregor[25] (31) in 1935 wrote a critical study of the application of the criteria for measurement to psychological phenomena, but Cohen and Nagel seem to have made the first important reference to the failure of psychological measurement to meet the necessary criteria. McGregor, basing his arguments on Cohen and Nagel, Bridgman and Campbell, reviews the criteria already discussed and applies them to psychological measurement.

His main conclusion is that there is no real difference between measurement in physics and in psychology. Operationally they are the same, as both depend ultimately upon the discrimination of difference. In this respect, he holds that from the operational point of view, the judgment of equality is "defined negatively in terms of inability to discriminate" (31). This is similar to the view expressed previously in this study. It was held that the equal judgment is in reality a judgment of no difference. McGregor's definition is, perhaps, a more accurate one, as it is true that the judgment of no difference is based upon the inability to discriminate a difference. However there is no contradiction between the two views.

McGregor's argument for the operational identity of the two types of measurement is sound enough if one goes no further than the basic operation of a "discrimination of difference." However, as has been seen, there are other operations in measurement. One difference between the two types was noted in the introduction.

The second point that McGregor makes is that such magnitudes as brilliance, chromatic saturation, loudness, weight, pressure, sweetness and pain are measurable in the ordinal sense. This, of course, the physicists will admit.

The third point that he makes is that equal sense distances are additive. First, he contends, it is necessary to obtain a series of equal sense distances. Then, "Demonstration of the first law[26] is easy,

[25] Working with E. G. Boring.

[26] I.e., $A + B > A'$ when $A = A'$ and $B > O$.

but demonstration of the second law[27] presents methodological difficulties, particularly when the 'sense distances' are not coterminous. Nevertheless, through the use of a method of substitution similar to that employed by the physicist in establishing a standard series of weights, we can demonstrate that the sum of a series of 'sense distances' is independent of the order of their addition.[28] This method of substitution enables us to define[29] the operation of addition much as it is defined for length in case I" (31). It is certainly not clear from this exactly what the definition of addition is.

H. M. Johnson (28) has published a criticism of "pseudomathematics" in psychology. He reviewed the logical requirements for measurement and discussed the measurement of brightness,[30] hue and attitude in the light of these criteria. The criticisms leveled against the attempts at measurement in psychology are extremely pertinent with one very important exception. His discussion of the additivity of brilliance brings up an extremely important point. To quote him, "Perceptible brightnesses are not identical with what the physicist calls the brightness or luminosities of surfaces. . . . It is imperfectly correlated with perceptible brightness within certain limits, but it is not identical with the latter.

"Consider a surface illuminated by two sources S_1 and S_2 in succession. When S_1 is used alone, the observer perceives a brightness of the surface which we may call B_1; when S_2 is used alone, he perceives a brightness B_2 on the same surface. Using the method of flicker photometry, or else the method of direct comparison, let us balance B_1 against a comparison-field B'_1, and also balance B_2 against another comparison-field B'_2. Now, expose the surface to both sources S_1 and S_2 at once. If we agree that in so doing, we add B_1 and B_2, then by the axiom of equals $B_1 + B_2 = B'_1 + B'_2$, and $B_1 + B'_2 = B'_1 + B_2$. But this is not generally true. Suppose, for example, that the sources which produced B_1 and B_2 respectively emitted only lithium light ($\lambda = 671$), while the sources that produced B'_1 and B'_2 respectively emitted only thallium light ($\lambda = 535$). Suppose, moreover, that $B_1 = B'_1$ is high, while $B_2 = B'_2$ is low. Then $B_1 + B'_2$ is the sum of a bright red and a dim olive green, while $B'_1 + B_2$ is the sum of a bright olive green and a dim red. Although observation yields the separate equations $B_1 = B'_1$, $B_2 = B'_2$, it is very likely to yield $B_1 + B_2 = B'_1 + B'_2$. The operations may not satisfy or even approximate the axiom of equals" (28).

You can't add w/ same diff dimensions

The first important fact to notice in Johnson's discussion is his clear recognition that the magnitude being scaled is a subjective magnitude. It is not the stimulus correlate, the physical magnitude, which is being measured.

The next important point to note is that which is contained in the sentence, "If we agree that in so doing we add B_1 and B_2, then by the axiom of equals $B_1 + B_2 = B'_1 + B'_2$. . . ." But why should one agree that the proposed combination of stimuli is an operation for the addition of the subjective phenomena? More generally it might be asked, why should one agree that *any* proposed method of combination should be accepted as the operation for the addition of *any* magnitude? As has been seen it is impossible to tell before experimentation whether any proposed operation

[27] I.e., IX, X, XI.

[28] The sentence "The sum of a series . . . of the order of their addition," is a verbal statement of the second law (IX, X, XI).

[29] Addition in the case of lengths was defined as the placing of the systems end to end.

[30] I.e., brilliance.

will meet the criteria for addition. It was also pointed out that this does not mean that the experimenter cannot make use of his past experience in choosing an operation which he feels has some chance of success. This is most probably what Johnson has done. On the basis of past experience with physical magnitudes it seems reasonable that this operation will add the magnitudes. In short, it meets the criteria of physical combination or physical juxtaposition. But it has also been shown that the criterion of physical juxtaposition has arisen because it is appropriate to physical measurement. Is there any reason to assume that the operation for addition of a psychological magnitude will be the same as that for the addition of a physical magnitude, especially when, as Johnson says, the physical magnitude is imperfectly correlated with the psychological? It might seem that the very existence of the imperfect correlation might lead one to expect that the operations would be different.

The operation for adding a subjective magnitude will be that operation that satisfies the necessary criteria for the subjective magnitude. This does not mean that the operations will or cannot be the same as those for the physical magnitude. It means that they cannot be the same when the relation between the physical and the subjective magnitudes is not linear.

It might be well to ask here how it can be known that the correlation between the physical magnitude and the subjective magnitude is not linear unless the subjective magnitude has first been measured. Actually it is not necessary to measure a subjective magnitude fundamentally in order to answer this question. All that it is necessary to do is to show that equal increments of physical magnitude do not correspond to equal sense distances. For example, if tones of 100 and 200 cycles are presented to an observer and he reports that the sense distance between 0 and 100 cycles is not equal to the sense distance between 200 and 400 cycles, the experimenter is justified in concluding that the two magnitudes are not linearly related.

It is necessary to go into this most important matter further. To give the example presented by Stevens (40): suppose that the observer was given a tone of 40 cycles and asked to find a tone equal to this in pitch. It is obvious that the observer would select another tone of 40 cycles (or very close to it). The experimenter now adds these two tones and presents a tone of 80 cycles and the observer is asked to find a tone equal to it in pitch. He will select a tone of 80 cycles. The experimenter now adds these tones and presents a tone of 160 cycles and asks the observer to find a tone equal to this tone in pitch, etc. Has the experimenter constructed a scale of pitch? It can be argued that it is a scale of pitch as the observer did not judge the physical correlate, the cycles per second, he judged the discriminable characteristic, pitch.

But as Stevens says, ". . . although at the outset we could conceivably choose any one of several sets of operations as defining the scale, that set will ultimately prove to be most satisfactory for a subjective scale when it leads to scale numbers bearing a reasonable relationship to the experience of the observer" (38). In other words, in the example used above, the numeral 1 might be assigned to the pitch of a tone of 40 cycles and the numeral 2 to the pitch of the tone of 80 cycles. The question is, do these numerals represent the subjective magnitude of pitch? Is the pitch of a tone of

80 cycles subjectively twice that of the pitch of 40 cycles; or, to put it more accurately in the terms of the operations which would be performed, is the pitch distance from 0 to 40 cycles equal to the pitch distance from 40 to 80 cycles? If it is not, the constructed scale does not correspond to the subjective magnitude of pitch.

It will be remembered that Campbell says that the operation of addition determines whether extensive magnitudes are the same. It might seem then that the operation of addition would determine whether the magnitude is subjective or physical. But the operation for addition will only distinguish between the two when the subjective magnitude is not linearly related to the physical magnitude. For example, if

__________is a line B, and
__________is a line B′, so that $B = B'$, and these two are added physically to produce another line C
_________________, is the resulting magnitude physical or subjective? By the criteria of identity of the additive operations it is necessary to declare that this is a physical magnitude.

But now suppose that zero physical magnitude is given the observer together with the physical magnitudes B and C, in this fashion:

A
O
____________B
__________________________C, and the observer is asked to judge whether the intervals $AB = BC$. The answer most probably will be yes. By this criterion, then, the magnitude is not only a physical magnitude but also a subjective magnitude. Furthermore, if the lines were presented in this fashion,

____________(B)
__________(B′)

and the observer was asked to add them subjectively and reproduce that line which was the sum of B and B′ it is highly probable that he would produce the same line that was obtained by the method of physical addition.

In short, the physical and subjective methods of addition would probably give the same results.

In other words, it seems that *if the relation between the physical and subjective magnitudes is perfectly linear, the operation of addition proper for one will also be proper for the other. If they are not so related, it will be necessary to find some other operation for addition in order to measure a subjective magnitude fundamentally*. However it must be noted that linearity over a very wide range can never be expected.

It is, of course, impossible to say in advance what these psychological operations for addition will be. One thing that seems certain is that they will not include physical juxtaposition except in the case where the subjective and physical magnitudes are linearly related. The experimenter will not add the physical correlates; the observer will add the two subjective magnitudes. They will be subjective in the sense that the equation of equal appearing intervals is subjective. The observer may be expected to make a judgment similar to that mentioned in Section E, 3, of an observer in one of the experiments reported below. He had, it will be remembered, two criteria for a ½ judgment; first, the equation of the intervals and, second, the check on this judgment by asking himself, "If I added B to B′ would they equal C?"

It is obvious that this sort of judgment will be rather difficult for many observers. If the experimenter is fortunate enough to find that the magnitude with which he is dealing is additive, and if

he also finds that the results obtained with the additive method correlate perfectly with the results obtained by the method of equal appearing intervals, he will, of course, be at liberty to abandon the more difficult method in favor of the easier. However it will always be necessary to show additivity and perfect correlation between the two methods before this is done.

As the method of equal appearing intervals (including the special case,—fractionation) is at the present time the only method for obtaining subjectively equal units, it would be well to review some of the important criticisms lodged against it.

For example, Guilford says: "But the reader must be reminded again of the still doubtful status of equated psychological intervals. There is the finding of Hevner that intervals among stronger stimuli are underestimated as compared with those among the weaker stimuli. Much earlier, Ament had found that, in a bisected interval, the higher of the two segments contained fewer j.n.d.'s than the lower. Whether the judgment of supraliminal differences can ever be brought into line with the judgments of liminal differences is hard to say" (21). Why should the two methods, based on different operations, yield the same results? Furthermore Guilford is assuming that *the* true scale for measuring a psychological magnitude is the j.n.d. scale so that the scale constructed from the method of equal appearing intervals must agree with it or be discarded. As we have seen, the method of equal appearing intervals is designed to give equal units. The operations used are those for obtaining, equality, i.e., the judgment of "difference" and "no difference" between the intervals concerned. In the method of just noticeable differences the equality of the j.d.n.'s is an assumption. There is no judgment of "difference" and "no difference" between two different j.n.d.'s. There is nothing in the operations for obtaining the j.n.d.'s that allows one to interpret them as equal. Without such an operational basis the statement of equality is meaningless. It may be true that they are equal, but this must be established by comparing the j.n.d. scale to a scale constructed upon a set of operations designed to give equal units.

Hevner (26) has found that the method of paired comparisons and the order of merit method gave the same results when used to measure goodness of handwriting. What she calls "the method of equal appearing intervals" did not give results comparable to the other two methods. Should this be surprising? As has been seen, there is no reason to be surprised if the use of different operations gives different results. In fact it is surprising if they give the same results.

Furthermore, Hevner did not actually use the method of equal appearing intervals. After all, the very essence of the method consists in making the intervals equal and Hevner did not instruct her subjects to make the intervals equal. Her instructions seem to have been copied from Thurstone and Chave (48), who also seem to have omitted the instruction to make the intervals equal. These experimenters, using what they call the method of equal appearing intervals, instructed their observers to arrange a group of statements, having to do with appreciation of the church, in the following manner: "You are given eleven slips with letters on them, A, B, C, D, E, F, G, H, I, J, K. Please arrange these before you in regular order. On slip A put those statements which you believe express the highest *appreciation* of the value of the church. On slip F put those

expressing a neutral position. On slip K put those which express the strongest *depreciation* of the church. On the rest of the slips arrange statements in accordance with the degree of appreciation or depreciation expressed in them. This means that when you are through sorting you will have eleven piles arranged in order of value-estimate from A, the highest, to K, the lowest" (48). There is no intimation given the observers that they are to make the interval between the statements placed in A and B equal to the interval between B and C, etc. All the instructions tell the subject to do is to arrange the samples in a given order, starting with A, the highest, and ending at K, the lowest, using F as the neutral or mid-point.

The observers may equate the intervals without being instructed to do so, but without specific instructions to the contrary there is no guarantee that they did not simply arrange the statements in a rank order of eleven steps. In fact, since that is all they were instructed to do, it is even probable that that is all they did do.

It is unfortunate that one must offer this criticism of this first important attempt to measure a discriminable characteristic which has no known stimulus correlate.

However Thurstone has used what he calls the method of equal appearing intervals under protest. He claims that "the ideal unit of measurement for the scale of attitudes is the standard deviation of the dispersion projected on the psychophysical scale of attitudes by a statement of opinion, chosen as a standard" (48). The logical fallacy of statistical scales has already been discussed.

Hevner takes a position somewhat similar to that of Thurstone. She says that there are "several facts that point to the superiority of the method of paired comparisons and the order of merit method over the method of equal appearing intervals" (26). To take these up in order:

1) The order of merit method and the method of paired comparisons give the same scale. The method of equal appearing intervals does not check with the other two.

2) There is no check on internal consistency in the method of equal appearing intervals. This, according to Hevner, is the most important criticism of the method.

3) In the method of equal appearing intervals the frequency distributions are skewed at the end of the scales. The medians for the ends are not as representative as the medians at the middle of the scale.

4) The j.n.d. or discriminal error, which is the fundamental psychophysical unit of measurement, is incorporated in both the order of merit method and the method of paired comparisons but is not involved in the method of equal appearing intervals.

To answer these objections in order:

1) The fact that two methods check one another does not mean that either one meets the necessary logical criteria for measurement. The fact that the method of equal appearing intervals does not check with the method of paired comparisons and the method of rank order does not mean that the method of equal appearing intervals is not acceptable; in fact from what we have seen it may, and probably does, mean that the other two are unacceptable as measures of subjective magnitude. The so-called "equality" obtained by the other methods results from statistical manipulations. The appropriateness of a scale must depend on the operations

used in constructing it, not on such mathematical manipulations.

2) Hevner has said that the method of equal appearing intervals lacks a check of internal consistency. But all of the logical criteria for measurement are checks of internal consistency. There is another check which has been applied by Gage to the consistency of loudness judgments. The method runs as follows:

Given the stimulus distance AE:

A B C D E

the observer is first instructed to equate the distance AC, CE. The observer is then presented with the stimulus distance AC and equates the distance AB, BC; and, likewise, he is given the stimulus distance CE and he equates CD, DE. The observer is then given the distance BD and he equates BC, CD. If there are no errors operating in the procedure the observer should "bisect" the distance BD at the original point for the "bisection" of AE, i.e., C. Incidentally Gage found that his final "bisection" was considerably higher than the initial "bisection." Newman, Stevens and Volkmann (32) used the same check on the judgment of loudness and, after making improvements in Gage's procedure, came to the conclusion that Gage's results were due to constant errors, all of which could be greatly minimized.

3) Hevner's third criticism is obviously not a criticism of the method of equal appearing intervals *as such,* but simply of her particular application of it.

4) The last criticism has already been answered in the discussion of the j.n.d. as a measuring device.

Several types of defense have been offered in support of the psychologists' position. One of the commonest may be summed up by quoting Bartlett and Craik, who were members of the above mentioned Committee of the British Association for the Advancement of Science.

"If all measurement must conform to the Laws of Measurement enunciated by Dr. Campbell, and, in particular, if the second law can only be satisfied by the physical juxtaposition of equal entities, then sensation-intensity cannot be measured. Yet this standard would have imposed a severe handicap in the early days of natural philosophy, and, maybe some sciences must still be allowed a greater latitude. The alternative would seem to be the coining of a new title for much that is called measurement.

"A scale built up by the addition of equal standard units is the ideal, but to say that measurement is possible only by such scales, would seem to be an unhelpful limitation of the meaning of the word" (Bartlett, R. J., 14).

"The Committee seems to me to have been facing two main points: (a) is sensation intensity measurable? (b) why should anyone want to make out that sensation intensity is measurable, and why should anyone want to measure it?

"The answer to the first of these questions must be sought by finding a definition of measurement which fits its use in other sciences, and then asking whether the facts obtained by psychological experiments enable the estimation of sensation magnitudes to be subsumed under this definition. It is important not to base the definition of measurement only on the most stringent instance, such as length; for 'measurement' is applied also to scales of temperature, density, time, etc. which fail to fulfill one or other of the conditions which are fulfilled by length. Thus, to insist that a quantity is measurable only if the operation of adding together two numerical quantities predicts the result of combining such quantities of the given

physical magnitude, would rule out the temperature scale quite as much as a sensation scale" (Craik, 14).

These statements give the impression that the whole point of the physicists' objections has been misunderstood. If one examines the statement of Guild's quoted on p. 6 it is true that he might seem to be forbidding the use of the term "measurement" to psychologists. But it seems to the author that the argument is not really over the use of the term, rather it is over the use of the meaning of the term as defined by the physicists. The logicians and physicists have appropriated the word "measurement" and have defined it in their own way. The definition is rigid and exact. The physicists say that psychologists may not use the term with this meaning because they have not demonstrated the relations necessary to give the term this rigid and exact meaning. Suppose the physicists and psychologists redefined the term "measurement" along the lines suggested by Craik above. They might arrive at some such definition as Scates (34, 35) *seems* to imply, namely, that measurement is anything that any intelligent scientist has ever called measurement. Then it would be necessary for the psychologists to use a new term to indicate what the physicists now call "measurement" or "fundamental measurement." Suppose they chose the word "scale." It is certain that the physicists would not now object to the use of the word "measurement" by the psychologists but would object strongly if they used the word "scale."

The author of this study suggests that Campbell's definition of measurement be adopted: the assignment of numerals to systems according to scientific laws. This definition would include all of the kinds of measurement discussed in this paper; the ordinal scale, Stevens' "intensive scale," the equal unit scale and both the A- and B-magnitudes. In order that this use of the term will not lead to confusion it is only necessary to prefix the correct adjective when one wants to speak of one of the specific kinds of measurement contained in the definition, such as "ordinal measurement," "equal unit measurement," etc.

When Bartlett says that the rigid standard imposed by the physicists would have been a severe handicap in the early days of physical science, it is difficult to see what he means. Certainly it is not a handicap to be unable to use the term measurement. The only handicap that the physicists could have suffered would have arisen from the inability to perform the necessary operations to demonstrate the required relations. But that does not mean that the early physicists did not continue to use what tools were at their disposal until better ones could be devised.

Likewise the physicists do not demand that psychologists immediately cease demonstrating those relations which they are able to demonstrate; but they do point out that the psychologists should not interpret their data on the basis of undemonstrated relations.

SECTION G. ZERO SUBJECTIVE MAGNITUDES

As was seen in the last part of Section B, a zero magnitude is associated with the proposition, if

$A + B = A'$ when $A = A'$,

B has the magnitude of zero.

Let A stand for the subjective weight correlated with a physical weight of 100 gr. and B stand for the subjective weight correlated with a stimulus increment of 3 gr.

If the two physical magnitudes are

added, the resulting psychological magnitude $A + B \not> A'$ when $A = A'$.

But if A is associated with a physical magnitude of 50 gr. and B still with one of 3 gr., the addition of the physical magnitudes will yield a psychological magnitude $A + B$ that is greater than A'. Thus B is O in one case and in the other case it is not zero.

Is this situation different from the situation existing in the realm of physical measurement? If the physicist, in weighing 5,000 lbs., adds 1 gr., then

$$5{,}000 \text{ lbs.} + 1 \text{ gr.} \not> 5{,}000 \text{ lbs.},$$

but if he is weighing 2 gr. on a sensitive scale and adds 1 gr., then

$$2 \text{ gr.} + 1 \text{ gr.} > 2 \text{ gr.}$$

In other words, zero is dependent upon the operations used for measurement. In this case the operations involve the use of scales of different sensitivity. If this is true, what then is the absolute zero? It would seem to be associated with a different system if one uses scales of different sensitivity. It seems logical that the absolute zero will be the magnitude of that system that fulfills

$$A + B \not> A' \text{ (when } A = A')$$

and the rest of the criteria for equality, on the most sensitive measuring device that has been devised.

In psychology the absolute threshold will be the magnitude that *will fulfill the above conditions for the whole range of any given subjective magnitude;* in other words, the *absolute threshold* would seem to be the logical zero magnitude for all discriminable characteristics.

SECTION H. MEASUREMENT WITHOUT PHYSICAL CORRELATES

When the physicist measures a fundamental magnitude the operations he performs do not depend in any way upon any other measurable magnitude. If they did he would not be in possession of an A-magnitude but of a B-magnitude. However it is necessary for him to be able to identify or reproduce the systems that he is measuring in respect of this given magnitude. If he is measuring weight, he must be able to tell one system from another. This could be done by marking the systems, say with letters of the alphabet, or by painting each one of them with a different color, red, green, purple, etc. Also, if he had previously measured their volume, he could identify them by their volume. But it is important to note that the operations he performs on the systems do not depend upon *volume as a measurable magnitude.* The volume merely serves as a means of identification and he could have used any other method of identification that was convenient.

Suppose that after he has measured the systems in respect of weight he plots weight against the identifying marks. That is, he plots weight against A, B, C, D, E, F, G, H, I, J, K. The plot might be something like that in Figure 3. But the question must be asked, why should the identifying marks on the abscissa be ordered in the way they are? Is it not equally valid to plot weight against the alphabet as arranged along the abscissa in Figure 4? The further question might be asked, why are the identifying marks equally spaced along the abscissa? Could not the plot be made as in Figure 5, with the alphabet scattered randomly along the abscissa? The answer is that both of these possibilities are equally valid because A, B, etc., have no meaning outside their use as identification marks. As identification marks they have no order and they are not necessarily equally distant from each other. The only reason for ordering and spacing A, B, etc., along the abscissa as they are in Figure 3 arises

from the fact that it has been found that the letters belong to systems that are ordered and equally spaced in respect of weight. In other words, the physicist could really be plotting weight against weight. Even though the units along the base line are letters of the alphabet, they obtain their relations to each other from the fact that they now represent systems that have been scaled with respect to weight. The plot of weight against identifying marks is meaningless.

Suppose, however, the physicist had used volume as his means of identification. Is the plot of weight against volume meaningless? There are two answers to this question. The plot is meaningless if volume is simply an identification mark, because if that is all it is there would be no reason for placing units of volume at equal distances along the abscissa any more than there was reason for placing the letters of the alphabet equally distant along the abscissa. But the plot is very meaningful if weight is plotted against volume *qua* volume. In this case weight, as the dependent variable, is plotted against another magnitude that is susceptible to fundamental measurement. The order and the equality of the units of volume in this case arise from the fact that volume has previously been measured fundamentally. In short, the physicist has plotted two fundamental magnitudes against each other and is in possession of a functional relation.

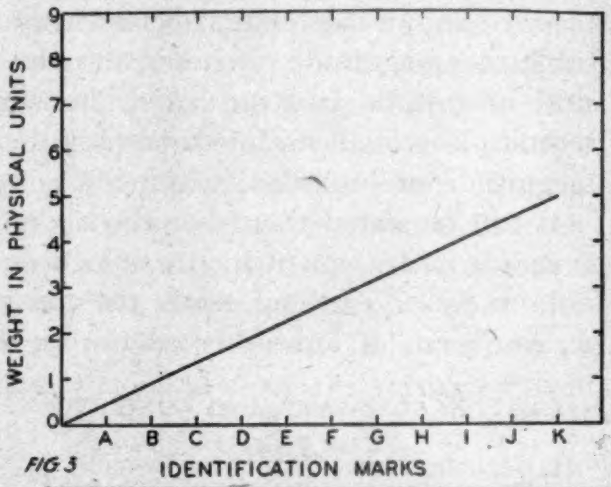

FIG 3

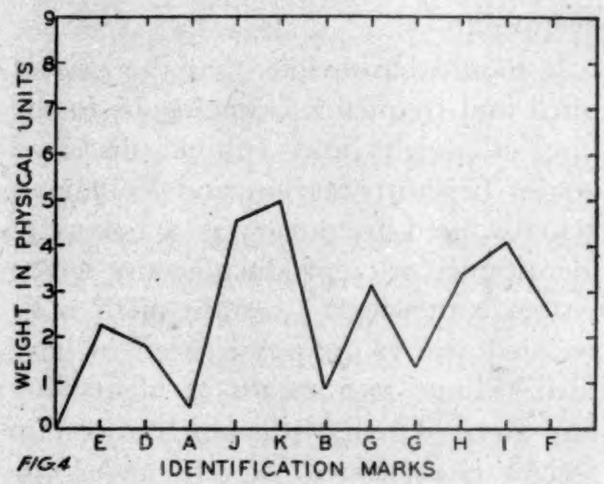

FIG 4

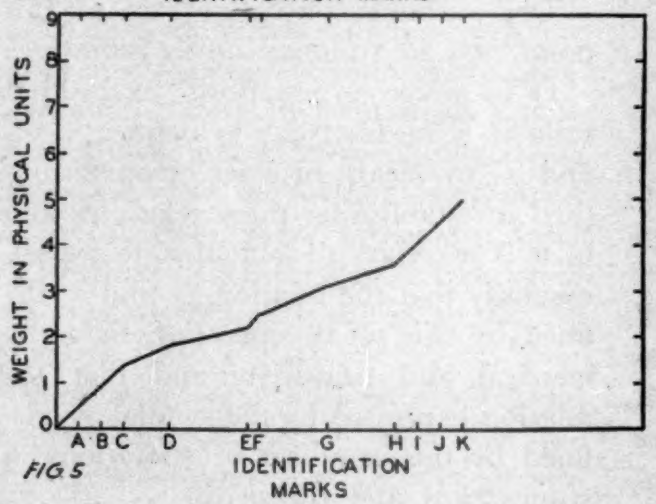

FIG 5

FIGS. 3-5. The "relation" between weight measured in physical units and identification marks (see text).

Turning to psychological magnitudes, consider the pitch function of Stevens and Volkmann, which was constructed after the fashion outlined in Section D.

First it is necessary to assume, solely for the purpose of this discussion, that the operations used by Stevens and Volkmann (42) have met all the criteria for measurement. Then the question may be asked, does the measurement of this psychological magnitude depend on the prior measurement of frequency. If it does, then pitch is not a fundamental magnitude.

It should be obvious that the case of pitch and frequency is analogous to the case of weight and volume discussed above. In short, Stevens and Volkmann actually used frequency as a means of identifying or reproducing any given system with which a certain pitch is associated, just as the physicist might have used volume as a means of identifying any given system with which a certain weight is associated. Stevens and Volkmann could have used any other convenient method of identifying pitch. For example, the identification might have been made by some arbitrary marks on the dial of the reproducing instrument. In short, pitch is measured independently of any other magnitude.

The pitch function happens to be a convenient graphic method for the assignment of numerals to different pitch magnitudes. *The fact that it also shows the relation of pitch to frequency is incidental so far as measurement is concerned.* When the magnitude of a number of identifiable pitches has been determined and the pitch function is constructed, the magnitude of intermediate pitches may be estimated by interpolation. However this must not be confused with measurement. The fact that the pitch magnitude associated with any frequency may be read from the pitch function does not mean that the pitch function is necessary for measurement. It simply means that, having measured a certain number of pitches fundamentally, the function can be used to give an *estimate,* albeit a reliable estimate, of the magnitude of other pitches. In the same way the weight-volume function might be used to estimate the weight associated with any given volume.

The pitch function may then be said to serve a double purpose. In the first place, it shows the relation between the subjective magnitude pitch and the physical magnitude frequency and, in the second place, it allows an *estimate* of the magnitude of intermediate pitches.

It can be stated then that the operations for measurement in psychology do not necessarily depend upon the prior measurement of any other magnitude.

SECTION J. SUMMARY AND CONCLUSIONS FOR PART I

The author's conclusions up to this point may be summed up as follows:

1) In order to establish an ordinal scale it is necessary a) to define $>$, $<$ and $=$ by means of a set of operations used in establishing these relations, and b) it is necessary to demonstrate experimentally that the relation $>$ and $<$ defined by this set of operations is asymmetrical and transitive; and that the relation expressed by the symbol $=$, defined by this same set of operations, is symmetrical and transitive.

In other words, with respect to $>$ and $<$ the following criteria must be met: if

$A > B$, then $B \not> A$ I (II)
$A > B$ and $B > C$, then $A > C$, II (II)

and with respect to $=$, the following criteria must be met:

$A \not> B$ and $A \not< B$ III

if

$A > C$ then $B > C$ IV
$A < C$ then $B < C$. V

2) When the criteria above have been met it is then possible to assign numerals

which may represent the relations demonstrated. Two types of numerals may be assigned, the conventional ordered numeral series, 1, 2, 3, 4, etc., or a series without a conventional order.

If the numerals with the conventional order are assigned to represent the systems in respect of the ordered magnitude, it will be convenient that they be assigned so that they may be interpreted conventionally, i.e., so that the conventional order of the numerals agrees with the demonstrated order between the systems. Since this is so, Campbell's rule for the assignment of numerals to an ordered magnitude must be followed. It must be borne in mind that the conventional order of the numerals adds nothing to the significance of the demonstrated relations.

If non-conventional numerals are assigned, the numerals will represent the order of the systems as in the case above. The chief difference is that when non-conventional numerals are assigned the person who uses the numerals must learn a new ordered numeral system, the order of which is determined by the order of the systems to which the numerals are assigned.

The assignment of non-conventional numerals emphasizes the fact that the numerals assigned to a group of systems do not create the relations between these systems. Experimentation demonstrates the relations, while numerals are merely used to represent previously demonstrated relations.

3) In order to construct an extensive scale it is necessary to define ( ) and + by a set of operations.

When this has been done, it is necessary to meet the following criteria: if

$A = A'$, and $B > 0$, then $A + B > A'$ VIII
$A + B = X$, then $B + A = X$ IX
$A = A'$ and $B = B'$, then $A + B = A' + B'$ X

and

$(A + B) + C = A' + (B' + C')$. XI

4) The above criteria are the only ones that it is necessary to meet in order to demonstrate additivity. It is not necessary to meet the criterion of physical juxtaposition, though physical juxtaposition may be necessary in order to meet VIII, IX, X, XI.

5) When the above criteria have been met numerals may be assigned to represent the systems.

As in the case of order, either a conventional or a non-conventional numeral series may be used. Whether the conventional or the non-conventional series is used, the assigned numerals obtain their meaning from the fact that certain relations have been demonstrated to exist between the systems to which they are assigned.

If non-conventional numerals are used and it is necessary to manipulate these numerals in order to predict the result of the actual manipulation of the systems, it would be necessary to invent a new arithmetic.

If, on the other hand, the conventional numeral series is used, it may be interpreted in the conventional manner, i.e., the numerals may be treated as if they represented objective *numbers* because the same relations have been shown to hold between the systems that hold between objective numbers. In short the ordinary, conventional arithmetic may be used to predict the results of actual manipulation of the systems. If the conventional numeral series is used and it is wished to interpret them conventionally, Campbell's rules for the assignment of numerals must be followed. If a non-conventional numeral series is used Campbell's rules for the assignment of non-conventional numerals may be followed.

6) The scaling methods based on differential sensitivity techniques do not meet the logical requirements for A-magnitudes nor the logical requirements for equal units. They do yield an ordinal scale.

7) The method of equal appearing intervals meets the requirements for an ordinal scale. This method also provides adequate operations for equating sense distances. The demonstration of equality of units does not demonstrate additivity. For this reason the numerals assigned to the magnitudes may not be interpreted as if they represented objective *numbers.*

8) The method of fractionation is a special case of the method of equal appearing intervals, except in so far as the observer gives himself the "additive instruction."

The half judgment or the bisection judgment cannot be operationally defined apart from 1) the equation of sense distances or 2) the addition of equal absolute stimulus intensities.

If the additive operation is not used, the same objections that are raised against the method of equal appearing intervals may be raised against the method of fractionation.

9) The statistical scaling methods present adequate operations for obtaining ordinal scales but not for extensive scales.

B-magnitudes may be constructed by the use of these methods. There are obvious precautions that must be observed in the interpretation of results obtained from the use of these methods.

10) The difference between a physical and a psychological magnitude will be determined by the operations for addition, except in the perhaps non-existent case where there is a linear relation between the two magnitudes. In those cases where the relation between the physical and subjective magnitudes is not linear, it is still possible that an adequate operation for addition of the psychological magnitudes may be found.

11) The operation for addition of the physical magnitude would often seem to involve physical juxtaposition, whereas the operations for the addition of the psychological magnitude would seem to involve subjective addition, i.e., a judgment of + without physical juxtaposition.

12) Psychologists may construct four types of scales:

1) Ordinal scales, i.e., those in which the relations of order have been demonstrated.

2) Intensive scales (as defined by Stevens), i.e., scales in which adjacent numerals are assigned by the adoption of some rule.

3) B-magnitudes, i.e., magnitudes that are scaled indirectly by some fundamental magnitude. There may be a postulated relation between the increments of the magnitude being scaled and the increments of the A-magnitude by which it is scaled. The B-magnitudes and the intensive magnitudes are very similar.

4) The scales constructed by the method of equal appearing intervals (or fractionation). These scales have both a demonstrated order and equal units. Additivity has not been demonstrated. The author disagrees with Stevens' contention that these magnitudes are examples of fundamental measurement. Hereafter in this study they will be called *equal unit scales.*

It is the author's bias that many psychological magnitudes will yield to the operations of fundamental measurement. To support this bias he can cite the following:

1) Introspectively the additive judgment seems possible.

2) Physical juxtaposition is not one of

the criteria for fundamental measurement.

3) The operations for constructing psychological magnitudes are independent of any other measurable magnitude.

4) It seems probable that perceived length (the discriminable characteristic that is correlated with the physical magnitude of *length*) will prove to be measurable by both physical and subjective operations for addition. If perceived length is measurable by a subjective operation for addition, the author sees no reason why other subjective magnitudes may not be measurable.

5) Campbell has expressed the opinion that the physical magnitudes that meet VIII ($A + B > A'$) will prove to be measurable fundamentally. He bases his conclusion on his experience with physical magnitudes and claims that he knows of only a few exceptions to this statement, e.g., intensity of x-rays.

If VIII is fulfilled, Campbell thinks that the fulfillment of the second law (i.e., IX, X, XI) will depend upon the perfection of apparatus and technique. In otherwords the essential criterion for additivity is $A + B > A'$, and the rest of the criteria will be fulfilled when the procedure is so perfected that constant errors introduced by apparatus, etc., are eliminated.

Certainly in psychology it is difficult to think of a case where the operation of subjective addition will not satisfy VIII. If an observer is presented with two lines,

A

B

and asked to reproduce a third that is equal to the sum of A and B, it is certain that the result will fulfill the criterion $A + B > A'$. Likewise if the observer is presented with two lights and asked to choose a third light whose brilliance is equal to the sum of the brilliance of the two, it is certain that the brilliance of the third light will fulfill VIII.

When these five statements are considered, fundamental measurement does not seem too far away.

It is true that no subjective magnitude has been measured fundamentally. The belief of the author that they may be so measured is an hypothesis. Only experimentation can give the answer. But it seems that the major objections of the physicists have been answered. *There are no a priori reasons why psychological magnitudes may not be measured fundamentally.* Measurement in psychology and physics are in no sense different. Physicists can measure when they can find the operations by which they may meet the necessary criteria; psychologists have but to do the same. They need not worry about mysterious differences between the meaning of measurement in the two sciences.

The author at one time believed that the equal appearing interval experiment led to fundamental measurement. He had not realized that the demonstration of equal units does not constitute a demonstration of additivity. The three experiments reported in this study use the method of fractionation. The results show that the three types of magnitudes here scaled are amenable to the method. They give further evidence that a physical correlate is not necessary for the measurement of psychological magnitudes. Unfortunately additivity has not been demonstrated, but equal unit scales have been constructed, and the author feels that the demonstration of addivity need be no longer delayed because of purely theoretical considerations.

Part III

Experimental: The Scaling of Visual Rate

SECTION A. THE PROBLEM

The problem in this experiment was to determine whether the perceived rate of the flash of a lamp is fundamentally measurable. It will be remembered that the author believed, during the course of these experiments, that both the method of fractionation and the method of equal appearing intervals were valid operations for the fundamental measurement of subjective magnitudes. As he has since changed his opinion, it might be better to rephrase the problem: The problem in this experiment is to determine whether an equal unit scale can be constructed for visual rate. Furthermore, if it is possible to construct an equal unit scale for visual rate, it is intended to find the relation between visual rate so scaled and its physical correlate.

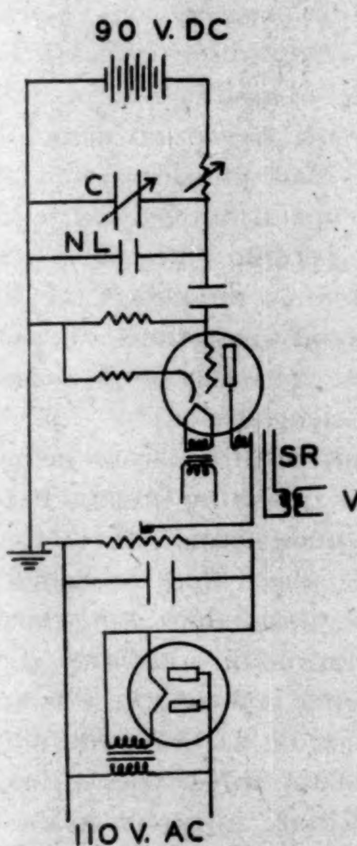

Fig. 6. Wiring diagram of the timer. C, the variable condenser; NL, neon lamp; SR, sensitive relay; V, the outlet to the lamp which served as the variable.

SECTION B. PROCEDURE

1) Apparatus

The current from two B-batteries is led through a variable condenser (*C*, Fig. 6).[31] When the condenser is fully charged it discharges through the neon lamp (*NL*, Fig. 6) which is connected in parallel. The amount of the condenser's capacitance will determine the length of time that it takes to charge—the higher the capacitance the longer the charging time. The rate at which the neon light will flash will be controlled, then, by the amount of capacitance in the variable condenser. A simplified diagram of the variable condenser system is given in Figure 7. The capacitance in this system ranged from 12 mfd to 0.10 mfd. This range of capacitance was capable of producing rates from 0.27 flashes per second to 12.3 flashes per second. The capacitance in this system could not be varied continuously. The change from 12 mfd to 0.10 mfd took place in 80 discrete steps. The average change in rate for each step was 0.15, though it was sometimes greater than this and sometimes less. The change in rate from one step to the next was very

[31] My thanks are due Mr. W. Rahm for this circuit.

frequently though not always below a noticeable difference.

The pulse set up by the condenser and neon lamp circuit is amplified by the vacuum tube circuit and used to drive a sensitive relay (*SR*, Fig. 6). The sensitive relay will make and break once for every flash of the neon light.

The current from a dry cell battery is run through the contact points of the relay to the primary of an inductorium. The current from the secondary of the inductorium is used to operate the variable stimulus (*V*, Fig. 6). Since a current will be produced in the secondary of the inductorium on both the make and break of the sensitive relay, a condenser is placed in parallel with the secondary of the inductorium. This condenser will absorb the "make current" but allow the "break current" to pass. In other words the variable stimulus will flash only on the break of the sensitive relay. The variable stimulus was produced by a 3 watt neon lamp, of the type in which the positive and negative poles are opposed to each other so that the observer could only see the positive pole through the hole in the shield. The lamp was mounted behind a shield in which there was a hole about 3/4 of an inch in diameter. The positive pole was toward the opening in the shield.

The observer on looking at the stimulus could see the brown shield with the grey of the positive pole of the neon lamp showing through the hole in the shield until the lamp flashed. When the lamp flashed the hole was filled with the orange-red glow of the lamp.

The purpose of the inductorium in the circuit was to keep the duration of the flash of the lamp constant as the rate of the flashing was varied by means of the condenser system.

The standard stimulus is controlled by a system that is almost identical. The chief difference between the circuits is that there are only 10 standard rates, so that the condenser system is much simpler than that for the variable stimulus.

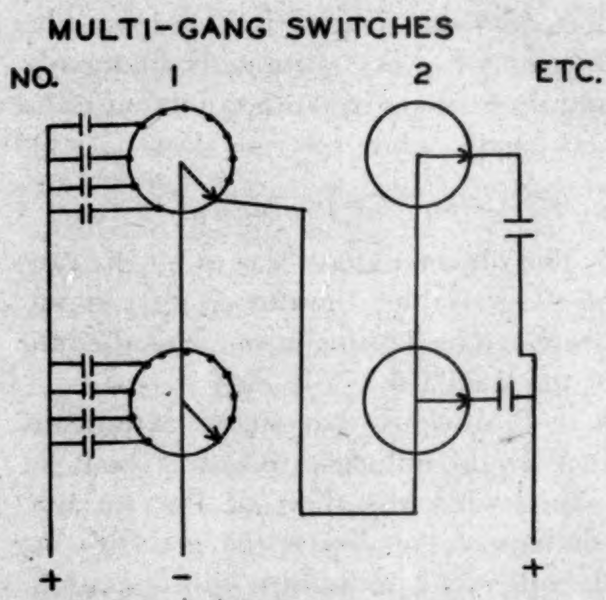

FIG. 7. Simplified wiring diagram of the variable condenser circuit.

It was possible to use the same amplifying system for the standard that was used for the variable, as the standard and variable stimuli were not presented simultaneously. The fact that they were not presented simultaneously necessitated a switching system so that the standard and variable stimuli could be presented alternately for fixed time intervals. The duration of the standard and variable was either 12 seconds each or 8 seconds each. The alternation of the two stimuli and the duration of their presentation was controlled by a Volkmann Timer (50). The circuit for controlling the standard stimulus and the Volkmann Timer circuit for controlling the alternation and the duration of the two stimuli are not shown in the diagram (Fig. 6).

The instrument was calibrated by connecting a signal marker in series with the primary of the inductorium and recording the pulses on a fast moving smoked drum. A time line was recorded on the drum by a high speed signal marker driven by a transformer which was connected to the commercial AC line. The instrument was calibrated frequently and the changes in rate were found to be very small.

2) Experimental Procedure

The observer's task was to set the rate of the variable stimulus so that it appeared to be flashing at one half the rate of the standard.

It is obvious that many extraneous cues would influence the observer if he manipulated the dials of the variable condenser. To obviate the necessity for the observer's performing this operation, the experimenter manipulated the dials at the demand of the observer. Thus, the observer would say "faster," if he thought the variable was flashing at a rate less than one half the rate of the standard; and "slower," if he thought the rate of the variable was more than one-half that of the standard. Since there was a fairly loud "click" each time one of the dials was moved one step, the observer soon learned to tell the experimenter by how many "clicks" the variable should be increased or decreased. In fact all observers, after the first, were informed of the possibility of using the clicks, and all of them readily adopted this method as it was definite, straightforward and easy.

In practice the observers would watch the standard, then the variable, and while the variable was still flashing, would say "Down five" or "Up three." Sometimes they made two adjustments of the variable during one presentation but they would usually wait until the standard and variable had been presented again. On the second and subsequent presentations of the variable they would continue instructing the experimenter as before, until they had reached a satisfactory ½ adjustment.

The standards were presented in random order. The original rate at which the variable was presented with any standard was changed for every presentation of that standard. Sometimes the variable was started at a rate that was very much faster or slower than one-half the standard, at other times it was started at a rate that was fairly close to the usual one-half judgment of the observer, and at still other times it was started at intermediate positions.

It was possible for the experimenter to produce the clicks without changing the adjustment or to produce a certain number of clicks that did not correspond to the number of steps that the dials had been moved. Both of these ruses were occasionally tried on all the observers but neither of them ever seemed to upset or affect the final judgment of the observer.

The observers were not given a set of formal instructions. The principle adopted was that all of the observers should fully understand the task and the operations they were to perform in accomplishing it. However the following points were always stressed for each observer.

1) The nature of the task.
2) The fact that the variable was always on the right.
3) The fact that they must instruct the experimenter to adjust the variable.
4) They were instructed to look at the center of the openings in which the lights flashed.
5) They were instructed not to count the flashes of either the standard or the variable.
6) They were instructed not to adopt any kind of rhythmical movements, such as

tapping with the hands, or feet, nodding the head, etc.

7) They were told that they might rest whenever they were tired.

The observers were seated at a distance of about two feet from the stimuli, which were set at an angle so that the shield formed an approximate right angle with the observer's line of regard. A daylight lamp (60 watts) shone on the shield and the two stimuli. The daylight lamp was protected so that it did not shine in the observer's face. The purpose of this lamp was to reduce the after-images of the stimuli which were distinctly noticeable in a completely darkened room. The lamp was effective in reducing the after-images to the extent that no observers seemed to notice them unless their attention was directed to them. Even then the observers reported that the images were very slight.

A trial series was given in which the observers were urged to ask questions concerning any portion of the task that might be disturbing them.

Each experimental period lasted approximately an hour, which usually included one rest period. Several times the sessions were somewhat longer. In this event another rest period was given.

There were five observers, who may be designated as *Du.*, *Co.*, *Lu.*, *Re.* and *Gr.*

SECTION C. RESULTS

1) General Results

It was noted above that the duration of both the standard and variable was either 12 seconds each or 8 seconds each. If the standard flashed for 12 seconds, the variable would flash for 12 seconds; if the standard flashed for 8 seconds, the variable would flash for 8 seconds. The experimental sessions were begun with the expectation that the 12 second duration would be used throughout the experiments.

It was found early in the first experimental session that the 12 second duration was too long for the faster rates. The observers became impatient because the lamps were flashing fast enough for the observers to obtain a good idea of the rate in a few seconds' time. However it was necessary to use the 12 second presentation time for the slower rates, because if this were not done, the lamps would not flash a sufficient number of times for the observers to obtain a good idea of the rate.

The question then arose: would the 4 second difference in the time of presentation make a difference in the observers' judgments? If the time did make a difference it would be necessary to keep the 12 second duration throughout the experiment. If it did not make any difference it would be possible to use the shorter duration of presentation for the faster rates and the longer duration of presentation for the slower rates. This would not only make the task of the observer simpler but would shorten the total time of experimentation. This was rather important as the observers could only make about 10 complete judgments an hour.

An attempt to answer the question of the influence of the duration of presentation on judgment, was made in the following way: the observer *Gr.* made five judgments of $1/2$ for each of 10 standard stimuli presented under the 12 second presentation and he also made five judgments of $1/2$ for the same 10 standard stimuli under the 8 second presentation time. The mean $1/2$ judgments were then compared. The differences between the means were tested by the t test and none of the ratios reached the 5 per cent level of significance. In

fact only one, that for the standard with a rate of 10.82 per second even reached the 10 per cent level of significance.

A summary of these results will be found in Table 1.

The interpretation of p in the last row of the table above is as follows: $p =$ the chances in 1.00 that a t as great or greater than that obtained would occur from a random sampling of a homogeneous population. For example the t value obtained for the difference between the means of the 1/2 judgments of the 13.98 standard is 1.51. From the table of probabilities under 8 d.f. (Fisher and Yates, 15) it is seen that this value lies between 0.2 and 0.1. In other words there are more than 10 but less than 20 chances in 100 that a t value as great or greater than this could occur in a random sampling of a homogeneous population. Since none of the obtained t values reach the 0.05 point, none of them may be regarded as significant. In short the chances are too great that they may have arisen by chance. This of course does not *prove* that the obtained means were drawn from a homogeneous population of means but it as least lends support to the view that the differences in the duration of the presentation of the stimuli may be disregarded. Accordingly it was decided to use the 12-second interval for the slower rates and the 8 second interval for the faster rates.

TABLE 1

Comparison of the means of the ½ judgments obtained under the conditions of 8 and 12 second presentation time.

Rate of standard in flashes per sec.	13.98	10.82	9.28	6.29	5.21	3.71	2.08	1.39	0.85	0.32
M of ½ judgments for 8 sec.	8.07	7.82	6.82	4.25	2.70	1.79	1.29	0.88	0.60	0.32
M of ½ judgments for 12 sec.	8.33	6.92	6.81	4.04	2.69	1.80	1.15	0.97	0.58	0.30
Difference between the M's	0.26	0.90	0.01	0.19	0.01	0.01	0.14	0.09	0.01	0.02
σ of the differences	0.172	0.477	0.286	0.196	0.238	0.165	0.103	0.049	0.028	0.014
t	1.51	1.89	0.03	0.97	0.04	0.06	1.36	1.84	0.35	1.38
p for 8 *d.f.* (Fisher and Yates, 15)	>0.10	>0.05	>0.90	>0.30	>0.90	>0.90	>0.20	>0.10	>0.70	>0.20

Observers *Gr.* and *Co.* each made ten 1/2 judgments for each of the ten standards. Observer *Lu.* made ten 1/2 judgments for all standards except the slowest, on which she made seven 1/2 judgments. Observer *Re.* made five judgments for all ten standards. Observer *Du.* made five judgments for all the standards except the slowest.

The experimenter started the experiments with the idea that at least ten 1/2 judgments would be needed to obtain consistent results. In Table 2 will be

TABLE 2

The M ½ judgments for each standard for the five observers, together with the σ_m, the σ dist., the σ_σ, and V

Observer *Gr.*

Rate of Standard.	13.98	10.82	9.28	6.29	5.21	3.71	2.08	1.39	0.85	0.53
M ½ judgment.	8.20	7.37	6.82	4.14	2.70	1.80	1.22	0.92	0.59	0.31
σ_m	0.12	0.32	0.20	0.14	0.17	0.12	0.07	0.03	0.02	0.01
σ dist.	0.38	1.06	0.64	0.44	0.53	0.37	0.23	0.11	0.06	0.03
σ_σ	0.08	0.24	0.14	0.10	0.12	0.08	0.05	0.02	0.01	0.01
V	4.7	1.4	9.3	10.5	19.6	20.3	17.3	11.8	10.7	10.3

Observer *Co.*

Rate of Standard.	12.90	9.84	8.11	5.45	4.53	3.16	1.75	1.16	0.71	0.43
M ½ judgment.	9.10	6.84	5.82	2.87	2.03	1.42	0.96	0.69	0.41	0.29
σ_m	0.35	0.36	0.35	0.21	0.18	0.16	0.09	0.02	0.02	0.01
σ dist.	1.12	1.14	1.11	0.67	0.56	0.52	0.28	0.09	0.07	0.04
σ_σ	0.25	0.26	0.25	0.15	0.12	0.12	0.06	0.02	0.01	0.01
V	12.3	16.7	19.0	23.3	27.3	36.5	29.7	12.9	16.6	14.2

Observer *Du.*

Rate of Standard.	14.00	10.40	9.1	6.25	5.26	3.70	2.04	1.41	0.61
M ½ judgment.	8.85	6.72	5.27	3.54	2.37	1.57	1.02	0.86	0.44
σ_m	0.19	0.67	0.51	0.37	0.21	0.05	0.01	0.04	0.03
σ dist.	0.43	1.51	1.14	0.84	0.46	0.11	0.02	0.08	0.06
σ_σ	0.14	0.48	0.36	0.27	0.15	0.03	0.01	0.03	0.02
V	4.9	22.4	21.6	23.7	20.0	6.8	1.5	9.9	13.9

Observer *Lu.*

Rate of standard.	13.98	10.82	9.28	6.29	5.21	3.71	2.08	1.39	0.85	0.53
M ½ judgment.	8.69	7.37	6.10	4.81	3.68	2.58	1.33	0.94	0.55	0.32
σ_m	0.27	0.16	0.26	0.18	0.16	0.10	0.08	0.05	0.02	0.02
σ dist.	0.84	0.52	0.83	0.57	0.51	0.33	0.26	0.14	0.07	0.04
σ_σ	0.19	0.12	0.19	0.13	0.11	0.07	0.06	0.03	0.02	0.01
V	9.6	7.1	13.1	11.9	13.9	12.8	19.7	15.4	13.5	12.8

Observer *Re.*

Rate of standard.	13.98	10.82	9.28	6.29	5.21	3.71	2.08	1.39	0.85	0.53
M ½ judgment.	8.69	7.89	6.58	4.48	2.87	1.95	1.24	0.84	0.55	0.28
σ_m	0.53	0.40	0.59	0.22	0.17	0.11	0.07	0.06	0.03	0.02
σ dist.	1.19	0.90	1.33	0.50	0.37	0.25	0.17	0.14	0.07	0.02
σ_σ	0.37	0.28	0.42	0.16	0.12	0.08	0.05	0.04	0.02	0.01
V	13.7	11.4	20.2	11.2	13.0	12.9	13.9	16.8	12.9	13.9

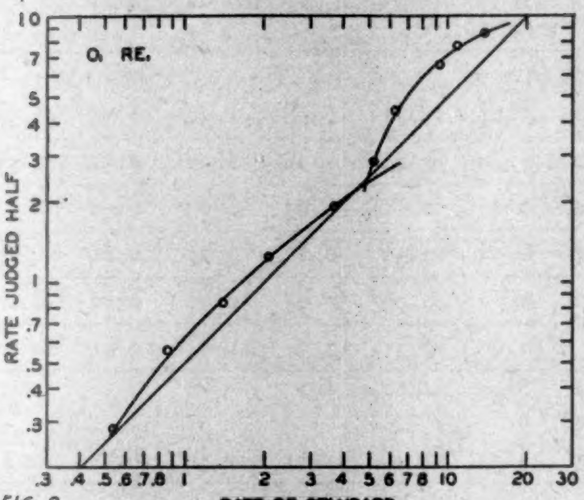

FIG. 8

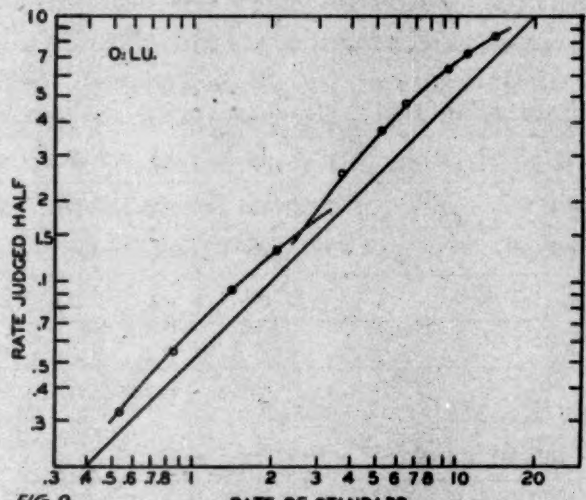

FIG. 9

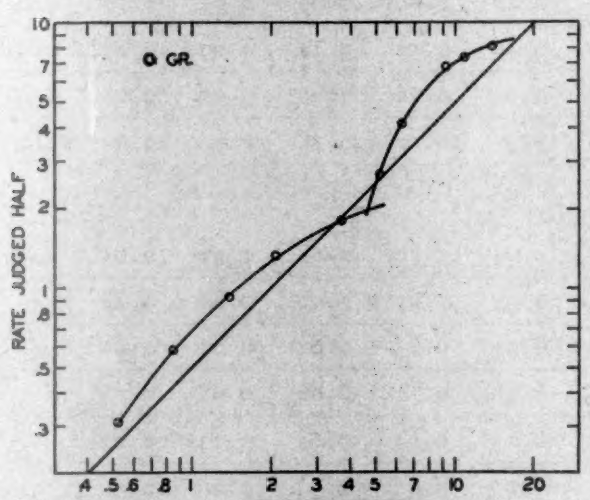

FIG. 10

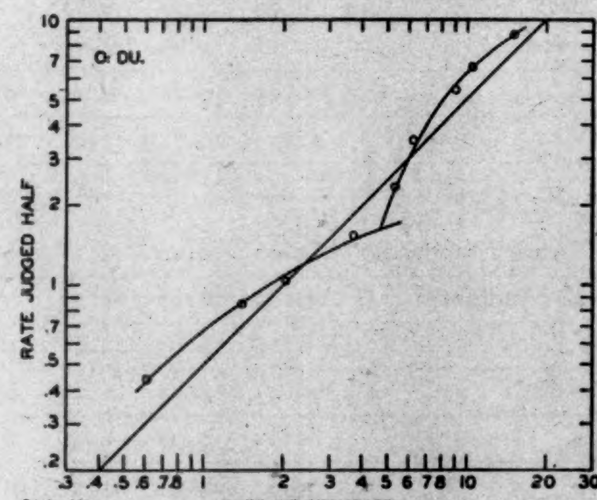

FIG. 11

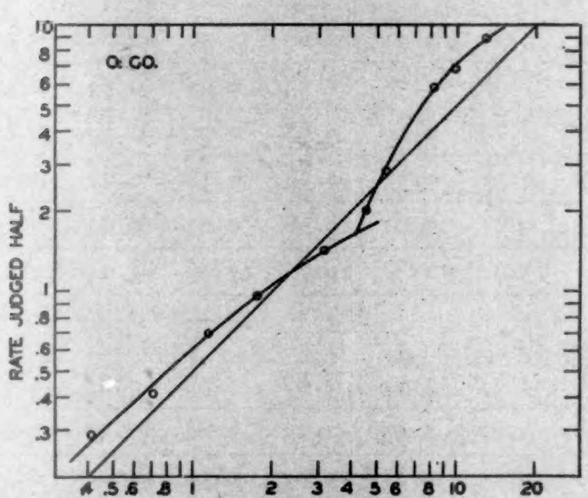

FIG. 12

FIGS. 8-12. The half judgments functions for visual rate for five observers (logarithmic coordinates). The coordinates represent objective rate, flashes per second.

found the mean 1/2 judgments for all the observers together with the σ_m for all the M[32] 1/2 judgments.

It should not only be noted that the σ_m for the 1/2 judgments based on an N of 10 are very small but also that the σ_m for the 1/2 judgments based on an N of 5 are also very small. While it is true that the σ_m's based on an N of 5 tend to be somewhat larger than those based on an N of 10, it seems legitimate to conclude that 5 judgments are sufficient to obtain a reliable result. When the very small N is considered it is obvious that the 1/2 judgment of visual rate is extremely consistent under the conditions of this experiment.

Table 2 presents the relative variability, V, and the σ of the distributions of 1/2 judgments for all observers, for all standards.

A graphical representation of the M 1/2 judgments for the five observers is shown in Figures 8-12. In each figure the M of the 1/2 judgments has been plotted against its standard. The plots have been made on log-log coordinates, with a diagonal straight line indicating the objectively correct half.

2) The Discontinuous Function

When the points had been plotted an attempt was made to fit a curve to the points by eye. It was immediately obvious that the points could be best fitted by two curves rather than by one. The data seemed to be discontinuous. While the discontinuity seems plain, the precise point at which the function breaks is not known. The point has been arbitrarily defined as the point of intersection of the two fitted curves.

The evidence in support of this hypothesis may be summed up as follows:

[32] Hereafter the symbol M will be used for the mean.

a) Taking any of the curves separately it seems as though two negatively accelerated curves would fit the data better than any other single curve. That is, unless the single curve had a sharp flexion point, in which case the sharp flexion point would in itself be some evidence of discontinuity. In Figure 13 will be found the data for *Re.* plotted on arithmetic coordinates (circles). Vertical lines, whose length represents 1 σ_M, have been drawn above and below their respective M's. A curve has been fitted to the data by the method of least squares. The equation for this curve is given by the polynomial,

$$y = -.31471 + .77586\,x - .00717\,x^2.$$

It will be seen that this curve misses seven of the obtained points by at least 1 σ_m and misses four of the obtained points by at least 3 σ_m.

In Figure 14 the same data have been plotted in the same way. In this case, however, two curves were fit to the data by the method of least squares. The equation for the lower curve is given by the polynomial,

$$y = -.0818 + .7491\,x - .05437\,x^2$$

and the equation for the upper curve by the polynomial,

$$y = -4.8853 + 1.83844\,x - .06202\,x^2.$$

It will be seen that the two curves lie within 1 σ_m of every obtained point.

The data for *Re.* was chosen for this demonstration because she had, on the average, the largest σ_m's.

This, of course, does not mean that the obtained points could not be fitted by a single curve merely by adding terms to the polynomial or by some other equation; it does show amply, if indeed it was not already obvious, that no continuous smooth curve without a sharp flexion point could adequately fit the data.

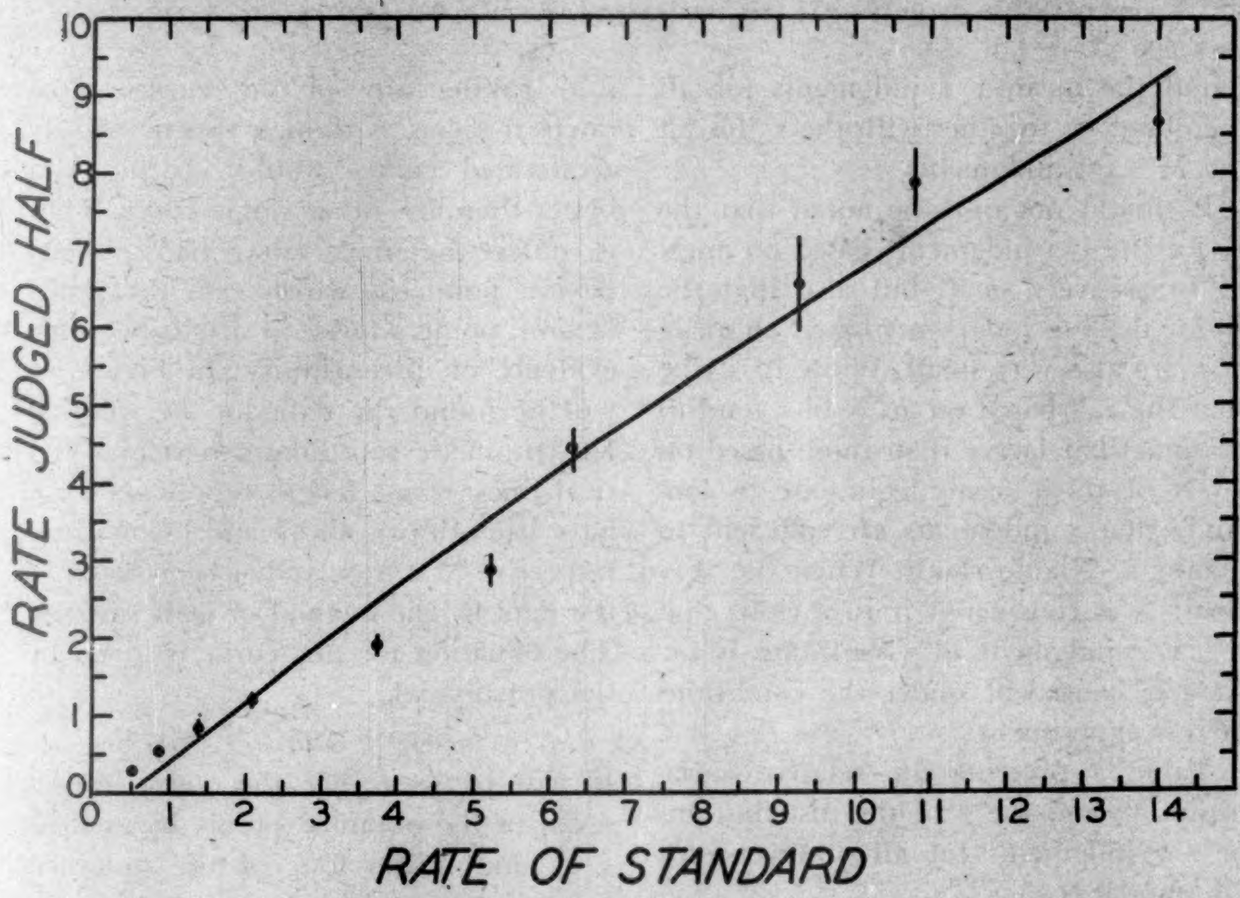

FIG. 13. *Re*'s data fitted with one curve (arithmetic coordinates). A vertical line above and below each point indicates $\pm 1\sigma_m$ (see text).

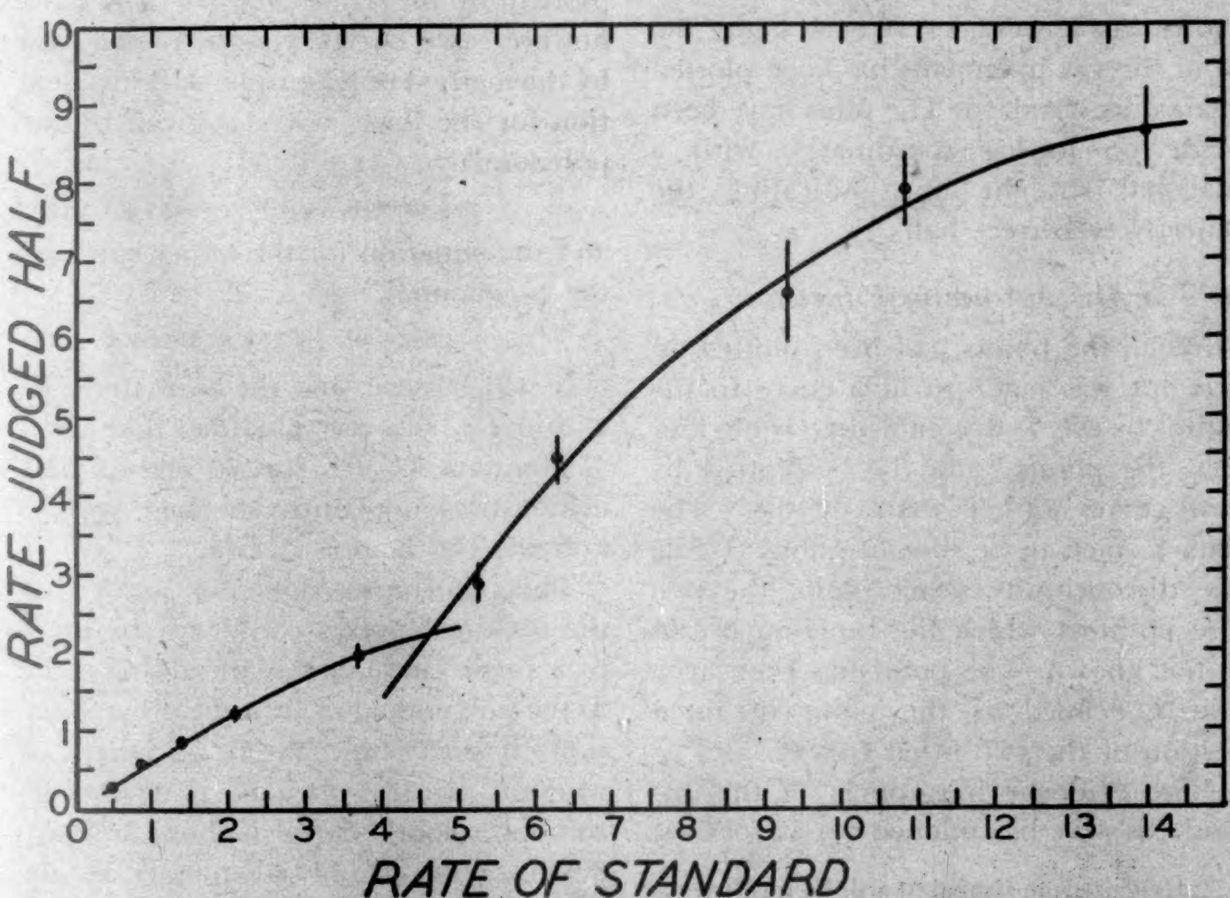

FIG. 14. *Re*'s data fitted with two curves (arithmetic coordinates). A vertical line above and below each point indicates $\pm 1\sigma_m$ (see text).

b) Further evidence for discontinuity is given by the fact that all the observers have curves of the same shape and all of them show the same evidence of discontinuity. If the data were truly continuous, it is reasonable to suppose that the function for at least one observer would appear to be continuous.

This argument becomes particularly potent when the small σ_m's are taken into account. A glance at the plotted σ_m for the data for *Re.* in Figure 13 or 14 should convince anyone that there is only a very small chance that the true means near the point of break could differ by much from those actually obtained.

c) Further evidence for discontinuity is given by the introspections of the observers. All of the observers found the slower rates harder to judge than the faster rates. Despite the consistency of judgments, all of the observers reported low confidence for all the judgments, but for the slower rates even this low confidence seemed to disappear and the observers often thought that they were simply guessing. The experimenter noted that a greater number of presentations of the standard and variable was necessary at the slower rates before the observer was able to come to a judgment.

All of this evidence points to the supposition that the task of the observer at the slower rates was different from the task at the faster rates. The only evidence of the nature of this difference so far is that the task is more difficult, that it takes a longer time and is accompanied by less certainty.

But all of the observers reported that they found the slower rates very "different" from the faster rates. When asked if they had adhered to the instructions and judged rate, and not the time between the flashes, all of them said that they had or thought that they had.

Observer *Co.*, however, reported that at one place in the series he seemed to base his "rate" judgment on the speed of the flashes and at the slower rates he seemed to base his "rate" judgments on "rhythm." It would appear that the observer was judging different discriminable characteristics of the same stimulus variable.

The experimenter then ran through the standard series alone and asked *Co.* to tell him when he came to the first standard that was judged on the basis of "speed." The experimenter started at the slower rates and presented each standard in succession. When he came to the standard of 4.53 per second *Co.* said that it was the first that was judged on the basis of speed. By referring to the graph for this observer, Figure 8, it can be seen that this is the first standard above the point of break. The experimenter then started at the faster rates and ran through the series and asked *Co.* to tell him when he came to the first standard that was judged on the basis of "rhythm." When the experimenter came to the standard of 3.16 per second *Co.* said that it was judged on the basis of rhythm. By referring to Figure 8 again it will be seen that this is the first standard below the point of break.

Furthermore observer *Du.* reported a very real difference in his perception of the slower and faster rates. When asked if the words "speed" and "rhythm" could be used to describe the difference between the two perceptions, the observer said that they could. He was not so sure that the word "rhythm" was as adequate to describe the slower rates as the word "speed" was to describe the faster ones. The experimenter presented the standards to this observer as he had for *Co.* Only a descending series was

given. The observer reported that the rates of 13.98, 10.82 and 9.28 could only be seen as "speeds," and that the rates of 6.29 and 5.21 could sometimes be seen as "rhythms" but were usually seen as speed and that below this point all of the standards were seen as rhythms. The break in *Du*'s function occurs between 5.21 and 3.71.

It will be noted that over a certain range the standard would appear as a "speed" and the variable as a "rhythm." For these observers, *Co.* and *Du.*, the average 1/2 judgments for the three fastest rates are seen as speeds. The average 1/2 judgments for the next two standards are seen as rhythms. In other words, the standards are seen as one thing and the variables as another. Since the break in the function occurs below this point (the point of discontinuity), both the standards and the rates that were judged 1/2 are seen as rhythms.

Co. claimed that when the standard was seen as a "speed" the comparison stimulus also *tended* to be seen as a speed.

The evidence above seems to point clearly to the fact that the observers are, in reality, judging two different characteristics of the flashing lamp stimulus. The observers are judging a "speed" when the lamp is flashing at a rate faster than 4.00+ per second; and they are judging "rhythms" when the lamp is flashing at a rate slower than 4.00+ per second.

The interpretation of the words "speed" and "rhythm" will be left to a later section.

d) Further evidence of discontinuity is offered by the relative variability of the 1/2 judgments, which changes characteristically at the point of discontinuity.

The coefficient of variability, V, for the 1/2 judgments has been plotted against the standard rates for all the observers. These plots are seen in Figures 15-19. The point where the 1/2 judgment plots break is indicated in these figures by a carat near the abscissa. The points have been plotted on semi-logarithmic coordinates.

It will be noted that all of these functions, except one, are strikingly similar in at least one respect. There is a large increase in relative variability at and around the point of break, followed by a sharp decrease.

This sharp increase in relative variability, followed by a decrease, could be due to several things. One possible cause would be the confusion caused the observer by the perception of the standard as a "speed" and the variable as a "rhythm." If this were the cause it might be expected that the highest variability would occur only in the 1/2 judgments of the two slowest "speed" standards. This, however, is not generally the case.

Another rather obvious possibility is that the standards near the point of break are sometimes seen as rhythm and sometimes as speed. Furthermore the variable for the two slowest "speed" standards might sometimes be seen as a "rhythm" and sometimes as a "speed." The introspections of *Du.* lend a little support to this theory. He noted, it will be remembered, that at least two of the standards could be seen both ways, although it was more natural for him to see them as "speeds."

It was said above that four of the graphs were similar, at least in that they showed a sharp rise in relative variability at the point of break and a drop after the point of break. If the V functions of the observers *Gr.*, *Co.* and *Lu*, are examined it will be seen that they are similar in showing a rather regular in-

crease in relative variability as the break is approached and a rather regular decrease after the point of break. *Du.*, while he shows an increase, does not show the same gradual increase, nor the same gradual decrease after the break. *Re.* shows neither the increase nor the decrease. The experimenter knows no reason why these observers do not show the same functions as the other observers, unless it be that the σ's of these two are based on an N of 5, whereas the σ's of the other observers are based on an N of 10. In short it is possible that these two differ from the "true" function because of the relatively less reliable σ's. There is a greater chance that their σ's are not the *true* σ's.

An examination of the $\sigma\sigma$'s lends little support to this explanation. Both *Du.* and *Re.* have relatively larger $\sigma\sigma$'s for the judgments for the faster rates but, on the other hand, the $\sigma\sigma$'s for the slower rates are approximately the same as those of the other observers.

If the true shape of the relative variability function is that shown by *Gr.*, *Co.* and *Lu.*, support would be given to the hypothesis that the stimuli at the point of break are sometimes seen as speeds and sometimes seen as rhythms. The further the stimulus is from the point of break the greater is the likelihood that it will always be seen as either one or the other and not as a mixture of both.

It is obvious that the support of the hypothesis of discontinuity given by the relative variability data is based on an assumption. The assumption is that the relative variability would not first increase, then decrease, in the manner shown in Figures 15-19 if the data were continuous. It seems fairly safe to make this assumption. While it is true that, in advance of knowledge, the relative variability function for a continuous 1/2 judgment function might be of any shape, it is improbable that it would show any sharp break.

3) The Construction of Magnitude Functions for Discontinuous Data

The construction of magnitude functions from discontinuous 1/2 judgment functions presents two types of special problems not found in the construction of the functions from continuous 1/2 judgment functions. The first problem is theoretical and the second is practical.

a) Theoretical. The first question that must be answered if the 1/2 judgment function is discontinuous is: should there be two magnitude functions or one? Campbell (6) claims that magnitudes are the same if the order generated between the systems is the same, regardless of the fact that the two orders are established by different operations. If the magnitudes are additive, they are the same magnitudes or are magnitudes of the same kind, if the operation for addition is the same. Campbell says that the magnitudes are the same or that they are magnitudes of the same kind because of some common property that determines the order in both cases.

It is necessary to examine this proposition of Campbell's further. Take for example the work of Taves (44) who has constructed a magnitude function for numerousness. If he had also constructed a magnitude function for density, with area constant, it is obvious that the order between the systems would be the same order as that for numerousness. Campbell would say that numerousness and density were the same magnitudes or magnitudes of the same kind. Yet Taves argues that density and numerousness are different magnitudes. He bases this argument on the fact that the 1/2 judgment function for density obtained

under the conditions of his experiment has a different shape from the 1/2 judgment function for numerousness. Is there a contradiction between Taves' position and that of Campbell?

It is true that these magnitudes are magnitudes of the same kind in the sense in which Campbell uses "same." Numerousness is a function of the number of stimulus dots; density (area constant) would also be a function of the number of stimulus dots. There is a common factor in the two series, although the subjective operations are different. In one case the observer is instructed to judge "numerousness" and, in the other, "density."

But the magnitudes are different in another sense. To construct a hypothetical example from physics, suppose

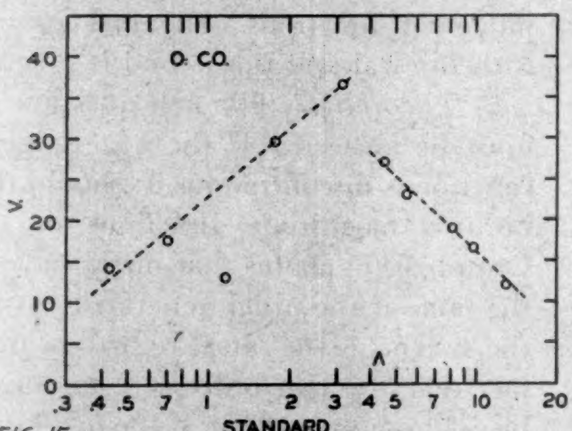

FIG. 15

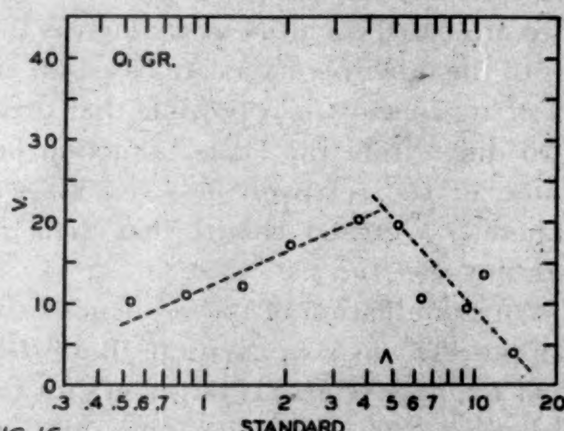

FIG. 16

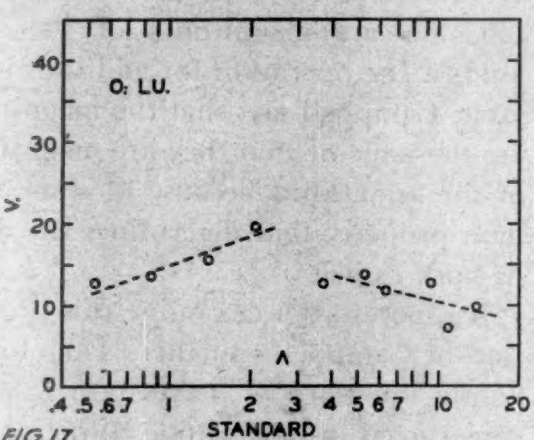

FIG. 17

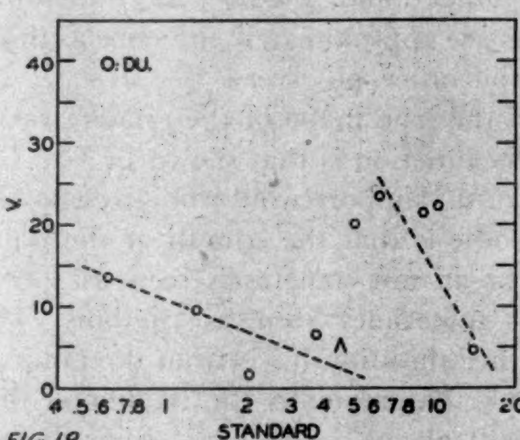

FIG. 18

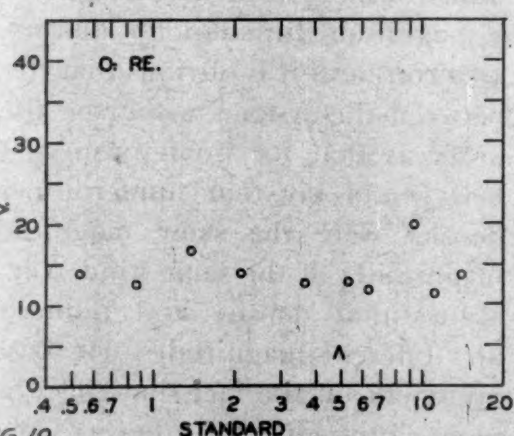

FIG. 19

FIGS. 15-19. Relative variability of the half judgments as a function of the rate of the standard (semi-logarithmic coordinates) for five observers. The abscissa represents objective rate, flashes per second, of the standard and the ordinate represents the coefficient of variability. The point of discontinuity as determined from the half judgment function is indicated by a carat.

that two B-magnitude scales have been constructed for temperature, one by means of the mercury thermometer and another by any other method. The order of the systems scaled by the two operations will be the same. However the functional relation between the two scales and a third measurable magnitude may be different. One might have a linear relation to some third magnitude and the other might have an exponential relation to the same magnitude; in other words there may not be a linear relation between the two B-magnitudes.

The magnitudes are the same in the sense that the order is the result of a common property—temperature. The fact that the relation between the two magnitudes is not linear points to the fact that they must be different in some other sense. The order obtained is not only a result of the common factor but also of the operations used to construct the scale. The sameness of the order can be said to be due to the common factor; the lack of linearity can be said to be due to characteristics of the operations.

To return to numerousness and density (area constant), in one case the observer is asked to judge the perceived number of dots and in the other case he is asked to judge the perceived number of dots per subjective unit area. The magnitudes are the same in Campbell's sense but they are different in the sense that the operation performed on the stimuli in the case of numerousness does not include the characteristics of subjective unit area.

This should give the clue for answering the question of whether there should be two magnitude functions or one magnitude function, when the 1/2 judgment data is discontinuous. There might be ways of telling whether discontinuity is evidence for two magnitudes:

1) It is possible to check the hypothesis that two magnitudes exist by changing the instructions. If the observers report that at one time they are judging "rhythm" and at the other "speed," it is possible to instruct them concerning the nature of the two characteristics and tell them to be sure to judge only the "rhythm" characteristic until it was absolutely essential to shift to "speed." Then the process would be reversed; the observer would be instructed to judge only the "speed" characteristic until it was absolutely necessary to change to "rhythm." If the change in instructions resulted in a shift of the point of break it could be presumed that there were two magnitudes.

2) Evidence of the existence of two characteristics could be gained from the verbal responses of the observers. If they describe two separate characteristics and the *verbal description checks with the obtained data* it may be presumed that there are two magnitudes.

b) Practical. If two magnitude functions are constructed there is no real practical problem. The functions may be constructed on the same coordinates or on different coordinates. There will be a separate point of origin for each curve.

If a single magnitude function is necessary the method is somewhat more complicated.

In Figure 20 the solid line represents some hypothetical 1/2 judgment data plotted on log-log coordinates. The form of the function and the type of discontinuity is the same as that for numerousness (Taves, 44). After assigning the numeral 1 to the physical stimulus value 1, successive points of the magnitude function may be plotted in the usual way. The resulting curves will be those shown by the solid lines in Figure 21. The type of break shown in Figure 20

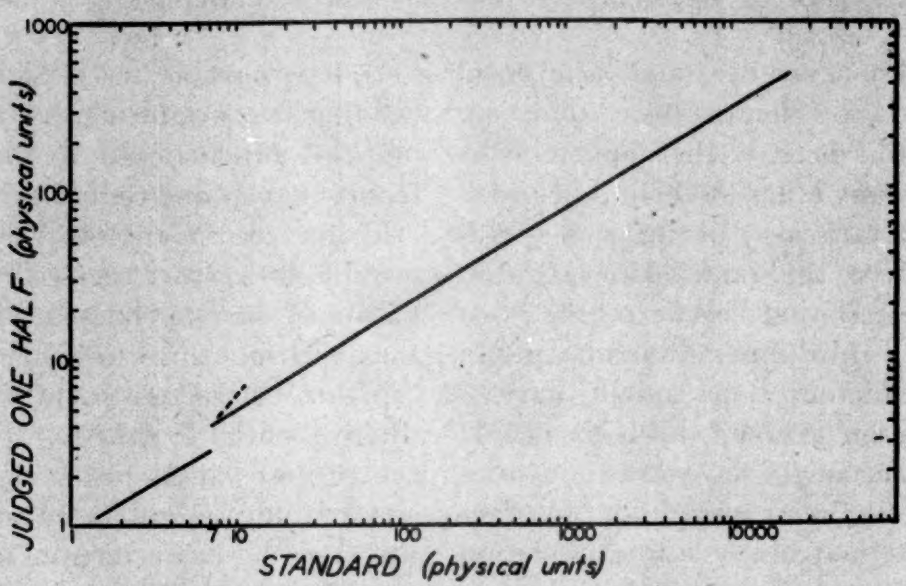

Fig. 20. Hypothetical half judgment function showing discontinuity (see text). The coordinates are logarithmic.

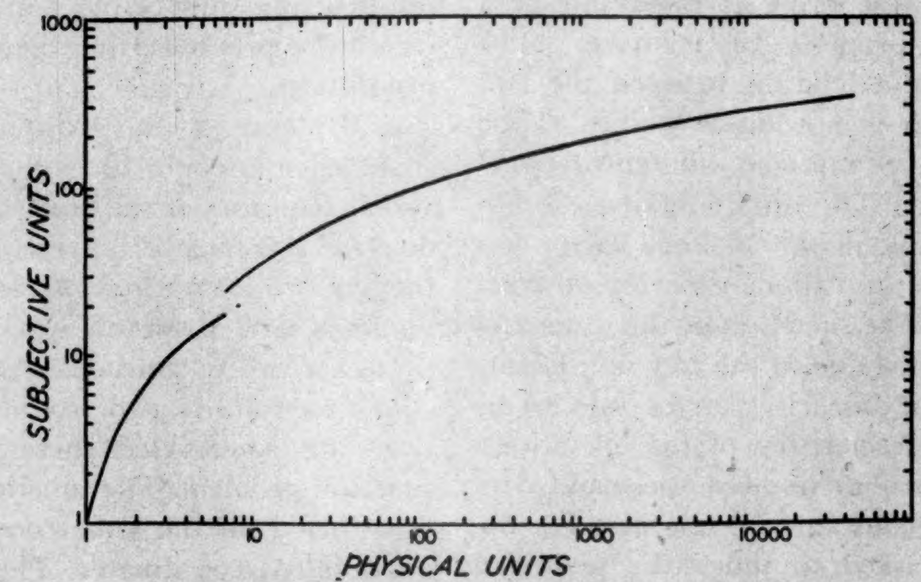

Fig. 21. Magnitude function showing discontinuity (see text). The coordinates are logarithmic.

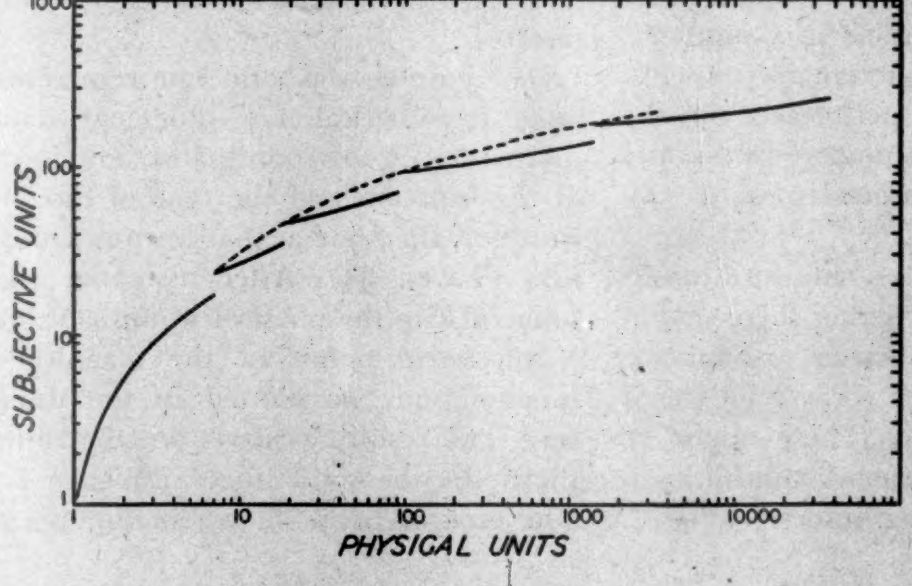

Fig. 22. Magnitude function showing multiple discontinuity (see text). The coordinates are logarithmic.

demands two values for the stimulus of 7 physical units. From Figure 20 it can be seen that 2.9 has been judged to be 1/2 of 7, and 4 has been judged to be 1/2 of 7. Therefore 7, which is the point of break, must have two subjective magnitudes. The question immediately arises: is this a practical possibility? If the observer judged 2.9 to be 1/2 of 7 half the time, and judged 4 to be 1/2 of 7 half the time, the mean 1/2 judgment for 7 would lie half way between 2.9 and 4. This is perfectly true. The only reason the break has been so drawn is to simplify the technique. Actually the break in the function will be in the nature of a steep curve between the values of, say, 6 and 8. However this does not change the basic principles that are being discussed.

In the discussion in Part I it was stated that the magnitude function served two purposes, first, to describe the functional relation between the subjective and objective magnitudes; second, to permit estimates of intermediate magnitudes by interpolation.

It was not mentioned that there is an objective check on this interpolated estimate. For example the stimulus of 8.5 physical units has a subjective magnitude of 28. The physical magnitude associated with 1/2 of this subjective magnitude is 5.0. Therefore 5.0 ought to have been judged to be 1/2 of 8.5. Turning to the 1/2 judgment function it is seen that 5.0 was not judged to be 1/2 of 8.5 but of 10.5.

It seems then that something is wrong with the magnitude function as constructed. It is in a sense not internally consistent when tested against the 1/2 judgment function. The first segment of the curve, that going from 1 to 7 physical units, is internally consistent. If the remainder of the magnitude function is actually plotted from the lower segment of the curve, i.e., from 1 to 7, rather than by simply drawing the best fitting curve between the obtained points, the following operations would be performed: Going to the 1/2 judgment function it is seen that 5.4 was judged to be 1/2 of 12 and, therefore, 12 should be assigned twice the subjective magnitude that was assigned to 5.4, and so on. The result will be 5 discontinuous functions. These functions are shown in Figure 22 by the solid curves. The magnitude function shown in Figure 21 has been drawn in in dotted lines. The amount by which the five discontinuous curves deviate from the magnitude function constructed in the usual manner is a measure of the lack of consistency of the function shown in Figure 21.

The function consisting of the five discontinuous curves truly reflects the relation between subjective magnitude and physical magnitude. The function is also internally consistent. Yet it would seem that there ought to be only one break in the magnitude function. There is only one break in the 1/2 judgment function. There ought to be no sudden jump in subjective magnitude at 18.5 physical units nor at 95 physical units, etc. The observer gives no indication of this in his actual judgments, as reflected in the 1/2 judgment function. On examination it is clear that the function in Figure 22 breaks as a multiple of 7, the original point of break. 7 was judged to be 1/2 of 18.5 and the function breaks at 18.5; 18.5 was judged to be 1/2 of 95 and the function breaks at 95; 95 was judged to be 1/2 of 1,430 and the function breaks at 1,430.

In other words the successive breaks were necessitated by the fact that there was a break at 7.

In the discussion of the discontinuity

of the visual rate function it was mentioned that over a certain range the standard would be perceived as a "speed" and the variable as a "rhythm." Over the rest of the range the standard and variable would be perceived as a "rhythm" and over another range they would both be perceived as a "speed."

If this fact is applied to the hypothetical data in Figure 20, it will be seen that the range in which the standard and variable are perceived differently must extend from 7 to 18.5. If this is so, is it reasonable that the curve from 7 to 18.5 is continuous with the curve from 18.5 to 4,000? It would seem that this is impossible. Since the effect of the break is to cause an increase in subjective magnitude, when the standard is a "speed" and the variable a "rhythm," the observer must select a variable that has a magnitude relatively greater than that which he selected when both the standard and variable lie in the upper range.

The final answer to the problem is now clear. The hypothetical data in Figure 20 could never have occurred. If there is one break in the 1/2 judgment there must be another break. The necessary shape of this middle segment is shown by the dotted line in Figure 20. It must be remembered that this shape is dependent upon the form of the break at 7, and that the form shown was used for illustrative purposes only. Actually it is reasonable to expect that the break will never be as sharp as the one shown. The result of the double break is to give a magnitude function that has only one discontinuity, that truly represents the relation between the subjective magnitude and the stimulus magnitude, and that is internally consistent.

This function is that shown in Figure 21 and by the dotted lines in Figure 22.

The evidence points to the fact that the break in the 1/2 judgment function for visual rate is due to the judgment of two different characteristics or, in other words, that there are two different magnitudes. Therefore two magnitude functions have been constructed. The origin of the first function is taken at 50 on the ordinate and 0.5 on the abscissa. That is to say, in constructing the magnitude function 50 subjective units were arbitrarily assigned to a rate of 0.5 per sec. The origin of the second function has been taken on the abscissa to be the point of break of the 1/2 judgment function; and on the ordinate to be the subjective magnitude assigned to the rhythm magnitude at the point of break. The points of origin of the second function, in terms of abscissa and ordinate, respectively, were for *Co.* 4.3 and 982; *Lu.* 2.65 and 700; *Du.* 4.09 and 620; *Re.* 4.8 and 800; *Gr.* 4.7 and 1,000. The two curves have been extrapolated for a short distance so that they intersect each other. This was done to emphasize the fact that they are functions for different magnitudes.

The magnitude functions for each of the observers are shown in Figures 23-27. The curves are plotted on log-log coordinates.

SECTION D. DISCUSSION OF RESULTS

Dunlap (10), using both visual and auditory stimuli and employing the method of constant stimuli, obtained the difference limen for both time and rate. Under one condition the observers judged the time between the stimuli and under the other they judged the rate of the stimuli. He found the difference limen for two standard stimuli for one of which the stimuli were presented at 0.232 second intervals (a rate of 4.31) and for the other the interval was 0.4355 seconds (a rate of 2.296).

He found that the sensitivity for rate

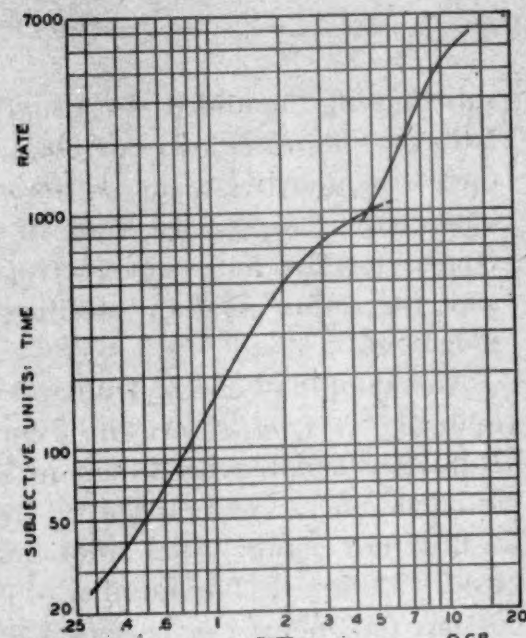

FIG. 23

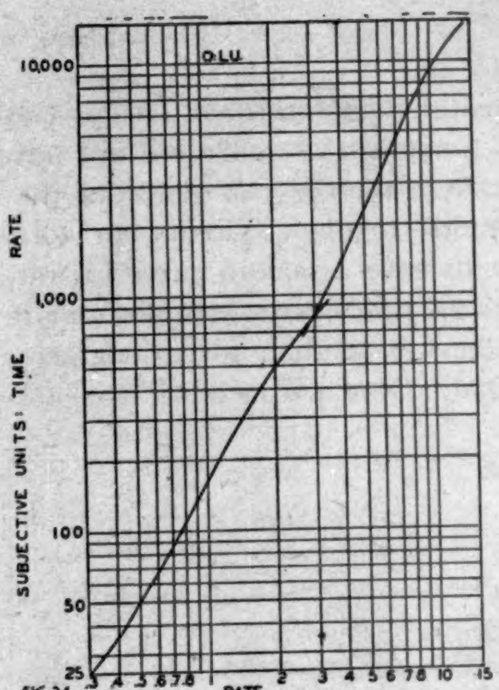

FIG. 24

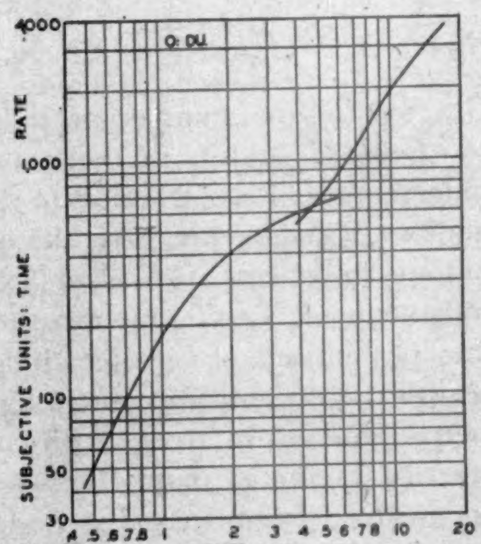

FIG. 25

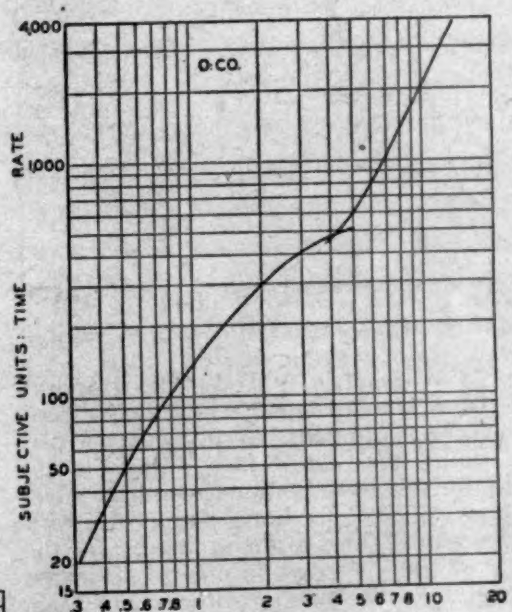

FIG. 26

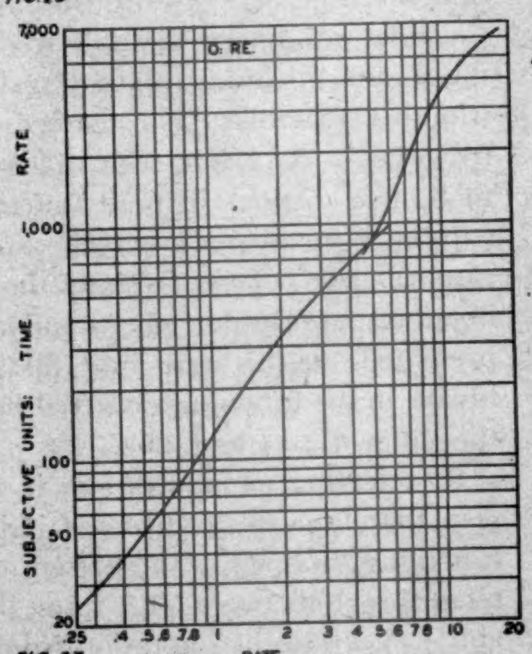

FIG. 27

FIGS. 23-27. The magnitude functions for visual rate for five observers (logarithmic coordinates). The abscissa represents the objective rate of the flashing lamp in flashes per second. The ordinate represents subjective time for the lower segment of the curve, and subjective rate for the upper segment of the curve.

was greater than for time. The psychometric functions for rate were not only steeper than for those of time but they were more regular. Twenty to forty judgments gave a smooth curve for rate whereas the time curves were not smooth even when more than forty judgments were made. Dunlap concludes from this

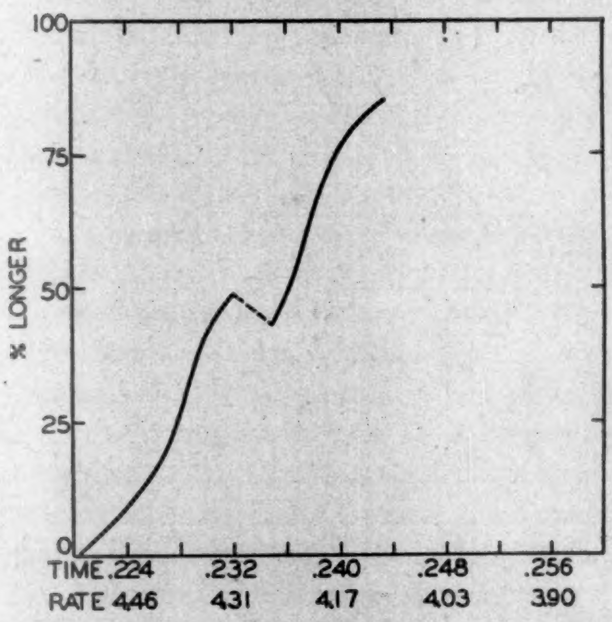

FIG. 28. Psychometric function for temporal intervals (from Dunlap, 100). The abscissa represents temporal intervals, in σ, between two auditory stimuli. A scale of rate, sounder clicks per second, for the corresponding temporal intervals has been added to facilitate comparison with the present study. The ordinate represents the percentage of times a given interval was judged to be longer than the standard interval of 232σ.

(p. 51) that, "These differences favor the supposition that the rate-judgment is not essentially a judgment of the interval between stimulations."

He then points out the possibility that the irregularities in the time functions are not due to poor judgment on the part of the observers. The irregularities in the curves are strikingly similar. They occur in 8 of the functions, 6 of which are for auditory stimuli separated by 0.232 sec. (rate of 4.31). Dunlap did not use visual time judgments so no data is available for the judgment of time intervals that are "bounded" by visual stimulations. The other two curves in which these irregularities occur are for stimuli separated by 0.4355 sec. (rate of 2.296). One is for the judgment of visual rate and the other for an auditory time judgment.

An example of one of Dunlap's curves (Dunlap 3, I) is shown in Figure 28. Dunlap calculated his limen in an unusual manner.

But the more usual plot may be readily obtained. The number of greater or less judgments can be found by solving simultaneously the equations,

$$G + L = N$$
$$G - L = N'.$$

G refers to the number of judgments of "Greater" and L to the number of judgments of "Less." N refers to the total number of judgments. N′ is the number of times judgments of "Greater" exceeded judgments of "Less." One can then plot the percentage of greater judgments against the variable and obtain a psychometric function in accord with common practice. Figure 28 shows the percentage of greater judgments (i.e., judgments that the time between the stimuli of the variable was longer than that between the standard stimuli) plotted against the stimulus variables. The nature of the irregularity is clearly seen. There seem to be two distinct sigmoid functions. It will be remembered that the standard rate was 4.31 per sec. In six of the curves obtained by Dunlap the standard was presented at this rate and all of the breaks in his functions occurred between the rates of 4.54 and 4.03.

By referring to Figures 8-12 it will be seen that four of the five 1/2 judgment functions obtained in the present experiment broke between the rates of 4.95 and 4.18. From the quotation from Dun-

lap given above he seems to have suspected that the "irregularity" of his psychometric functions for time might be due to something more than unreliable discrimination. The analysis of the data obtained by Dunlap for the time judgment gives added support to the contention that there is normally a change from the discriminable characteristic "speed" to that of "rhythm" (the meanings of these two words are expanded below) in the neighborhood of 4.00-5.00 stimulus presentations per sec.

In the section above introspective evidence was offered to support the hypothesis that the judgments of the faster and slower rates were based on different discriminable characteristics.

These characteristics were named "rhythm" and "speed" by the observer *Co.* It is now necessary to attempt to interpret these two terms. One thing was certain from the beginning "rhythm" was not used by *Co.* with the usual connotations. When *Re., Du., Co.* and *Gr.* were asked to verbalize the two characteristics more fully, they replied in substance that: "When the light flashes at a rapid rate it seems to be there all the time. When the light is flashing at a slower rate it is not there all the time. There are blank intervals bounded by the flashes."

Tinker (49) has stated the difference between rapid and slow rates as follows, "The rate, however, must not be too slow, for in that case the perception of the interval betwen the qualities or events tends to intrude and to become the characteristic feature of the perception, and the succession as such becomes difficult to perceive."

The experimenter believes that the so-called "speed" judgment was really a judgment of "speed" or "rate" but that the so called "rhythm" judgment was a judgment of the interval between the flashes. The question now arises, why do the observers cease judging on the basis of rate at the particular point where they do?

Dunlap (10) has reported that when flashes of light are of equal length the smallest perceptible interval between the flashes ranges from 0.004 sec. to 0.0198 sec. The durations of the flashes of light ranged from 0.0198 to 0.0729. The shorter time intervals between the flashes were associated with the longer durations. Dunlap notes that intensity is an important variable in this determination.

These intervals were extremely small. If the observers are able to perceive time intervals as small as this, why do the observers in the present experiment change from a time to a rate judgment when the durations are approximately 0.222? The experimenter believes that the answer has been given by James (27): "To be conscious of a time interval at all is one thing; to tell whether it be shorter or longer than another interval is a different thing." In short when the interval reaches a certain duration any further decrease is imperceptible. The subjective magnitude of time becomes asymptotic to the absolute threshold and the subjective magnitude of rate will become asymptotic to the upper limit. As the flashes increase and approach the fusion point the subjective magnitude of rate should be negatively accelerated. It will be asymptotic to the fusion point.

This does not mean that the observers in the present experiment judged on the basis of time as long as they could. It is the opinion of the experimenter that they judged on the basis of rate as long as it was easy to do this and, conversely, that they judged on the basis of time as long as it was easy to do that. The hypothesis is suggested that at a certain point

it is equally difficult for the observer to judge on either basis. It is possible that at this point variability should be at its highest. This does not mean that the point of break in the 1/2 judgment functions marks the point above which the observers *cannot* discriminate differences in time and below which the observer *cannot* discriminate difference in rate. It is the point at which it is no longer very easy for the observer to continue to discriminate either time or rate.

There is still another possible explanation of the discontinuity. There is some change in brilliance as a function of rate. It is possible that the integrated intensity of the flashes begins to produce an effect on behavior in the neighborhood of 4.5 per sec. At the present time the author can do no more than present the possibility as he has no direct experimental evidence to either refute or substantiate it.

SECTION E. SUMMARY

1) The problem was to construct an equal unit scale for visual rate (the perceived rate of the flash of a lamp) by the method of fractionation and to determine the relation of visual rate so scaled to the stimulus magnitude.

2) The 1/2 judgment function showed marked evidences of discontinuity. The hypothesis that the 1/2 judgment function for visual rate is discontinuous is supported by a curve fitting technique, by the fact that all the observers show the break, by the break in the relative variability functions and by the introspections of the observers.

3) The discontinuity is attributed to the possibility that the observers may have been judging time at the slower rates and "speed" or rate at the faster rates.

4) A solution was offered for the theoretical and practical difficulties inherent in the construction of magnitude functions for discontinuous data.

5) The extremely small variability of the 1/2 judgments and the consistency between the observers in respect of the shapes of the 1/2 judgment curves lead to the belief that the results are reliable.

6) An equal unit scale which would meet the criteria for ordinal scales and for equal units has been successfully constructed for visual rate.

Part IV

Experimental: The Scaling of Subjective Difficulty of Digit Series

SECTION A. THE PROBLEM

The problem is to discover whether an equal unit scale can be constructed for the subjective difficulty of memorizing and recalling digit series. If a scale for subjective difficulty for digit series can be constructed, it will be possible to determine the relation between the subjective difficulty and the number of digits in the series. The scale could also be compared to the objectively determined incorrectness of the reproduced digits and to the subjectively determined incorrectness of the reproduced digits. The objectively determined incorrectness is defined as the percentage of the series of any given length that is reproduced incorrectly and the subjective incorrectness is defined as the percentage of the series of any given length that the observer thinks he reproduces incorrectly.

SECTION B. DISCUSSION OF THE PROBLEM

1) Digit Span

The digit span has been used as a test of memory for some time. It was included in the Binet-Simon tests (2). Perhaps digit span finds its greatest practical use in present day psychology in various well known intelligence tests. In the Stanford Revision of the Binet-Simon Tests (45) a series of 3 digits is placed at the three year level; four digits at year four; five digits at year seven; six digits at year ten; seven digits at year fourteen; and eight digits at the superior adult level.

In the Revised Stanford-Binet Scale (46) two digits are placed at two years and six months; three digits at three years; four digits at four years and six months; five digits at seven years; six digits at ten years; eight digits at Superior Adult II and nine digits at Superior Adult III.

The practice in the Stanford-Binet Tests is to count the test as passed if one out of three of the digit series at any given level is correct.

Kuhlmann's Tests of Mental Development include a digit series of two at two years two months and one of five at five years one month. The directions are not the same for the two year-levels.

There have been several determinations of auditory memory span for digits at the adult level. Carothers (8) found an auditory span of 7.53 ± 0.53 for a group of women college students ($N = 200$). Garrett (18), using a group of 158 male college students, obtained a span of 8.4. Martin and Fernberger (30) report that college students are able to increase their digit span about 20 per cent with practice.

The method of constant stimuli provides a more accurate technique for obtaining the memory span for digits than that of mental test technique. The digit span obtained by the mental test technique does not have the same meaning as the digit span obtained by the method of constant stimuli. For example, on the Stanford-Binet the test is passed if the observer reproduces one series correctly out of three. All that can be said of the result is that the observer was at the time of examination at least capable of reproducing one series of a certain

length, when given three trials. When the span is computed by the method of constant stimuli a large number of series of different length is given to the observer and the percentage of correct (or incorrect) reproductions is plotted against the number of digits in the series. The median correct series and the P.E. can be obtained graphically; and the M and σ can be obtained by any one of several methods, (Woodworth 52, p. 402 ff.).

2) Subjective Difficulty

The usual method of measuring the difficulty of a mental test item is in terms of the frequency of correct response. A difficult item is defined as an item that is passed by relatively few persons. An easy item is defined as one that is passed by a relatively large number of persons. The results of the measurement of difficulty defined in this manner are expressed in percentages or σ units or *t* scores or other measures depending upon the frequency of passing or failing the item. It has been suggested (51) that "difficulty" is a poor name for this phenomenon. For example, "We may, then, conclude that the concept of difficulty (in its objective aspects) should for purposes of precise statement be renamed: and we suggest some such phrase as 'incidence of successful performance' (under specified conditions) as relatively free from the possibility of misinterpretation through assumption of more or less magical properties to units of the task itself. . . ." Difficulty defined in terms of frequency of failure will be referred to hereafter as *objective difficulty.*

There is, of course, another kind of difficulty which is part of the experience of the observer. If an observer is asked if an item is difficult he is able to give a rather definite reply. If asked to explain the basis of his judgment he may give a variety of material. A list of some of these introspections follows; the list is based partly on the introspections of the observers in this and the next experiment:

1) Estimation of how difficult the task would be for a large number of people ("difficult" being defined objectively).

2) Estimate of the correctness of the observer's own answer.

3) Confidence in final answer *after* the task is completed.

4) Confidence that a correct reply will be obtained *during* the course of the reproduction.

5) Lack of familiarity with the type of task (the task may be subjectively difficult because the observer has never had any experience with the material).

6) Length of time to solve the problem.

7) Complication of the problem (i.e., the problem may be long and intricate but objectively easy).

8) Feelings of strain and effort.

9) Feelings of indecision.

It can be seen that many of the above criteria for subjective difficulty overlap. It can also be seen that some of them are based directly on an estimate of the correctness or incorrectness of the observers' final reply. The experimenter did not wish to define subjective difficulty on the basis of the observers' direct estimate of the correctness of the final answer. He wished to scale the difficulty of *doing* the item. In other words the observers were asked to use as a criterion the "difficulty" they experienced in reaching a solution no matter whether they thought that that solution was correct or not.

Beyond this and a few cautions against false criteria no definition of subjective

difficulty was attempted. The experimenter did not wish to adopt a rigid definition of difficulty and then warn the observers not to judge on the basis of this or that criterion. He felt that if this was done there might be no criteria left on which the observers could base their judgments. Actually the observers reported approximately the same criteria, numbers 4, 6, 7, 8 and 9 above. But these criteria were not equally effective for all the observers. An observer would report, for example, that he used 4, 6, 7 and 8 but that 6 was the most important determiner of the judgment. It is the opinion of the experimenter that all of these so called "criteria" are really different discriminable characteristics that are closely related to each other. They are magnitudes of the same kind. When the observer is asked to make a judgment of "subjective difficulty" he actually judges one or more of these characteristics. Some of these have been excluded by the instructions not to use "correctness" as a basis for judgment. The relative importance of certain criteria will not only be a function of individual differences but also of the material.

Farmer (12) found that there was more subjective difficulty for objectively difficult tasks than for objectively easy tasks, even when the observer did not know the objective difficulty of the task. Hertzman (25) found rank order correlations ranging from 0.50 to 0.86 between confidence ratings and the objective difficulty of memory items that were correctly matched. Hertzman defines subjective difficulty by "confidence" (p. 114).

SECTION C. PROCEDURE

1) The observers were given a sheet of paper made up as follows:

Standard	Variable	Judgment	Errors in repro.
1 ———	———	G ½ L	—— ——

There were enough of these spaces for all the digit series given in a single experimental session. The observer was given a copy of the following instructions, which he was asked to read while the experimenter read them to him:

I am going to read a series of digits to you and when I have finished, I want you to write it from memory in this space (space for standard). I am then going to read another series to you and when I have finished, I want you to write *it* in *this* space (space for variable). Then I want you to judge whether the second series of digits was half as hard for you to remember as the first, or whether it was more or less than half as hard. If the second series seemed half as hard, circle ½, if it seemed more than half as hard as the first, circle G, and if the second series seemed less than half as hard as the first, circle L. When you have finished making this judgment of the relative difficulty of the second series, will you put an X in this column if you think you have made an error in reproducing the first series; and put an X in this column if you think you have made an error in reproducing the second series.

Don't worry about mistakes or what you think are mistakes as I shall give you quite a few series of digits that are too long for you to remember. First, try as hard as you can to learn each series, then make the judgment of difficulty and after that register your impression as to whether each of the series were reproduced correctly. Do each thing in the order in which I have given it.

N.B. You are to judge whether the second series is *half* as difficult to learn as the first, or whether it is *more than half* as difficult or *less than half* as difficult, NOT whether the second series is *as difficult* as the first or *more difficult* or *less difficult.*

To look at it another way, the nearer the second series approaches the first in difficulty the more likely it is to be more than half as hard, and the easier the second series is in reference to the first the more likely it is to be less than half as hard.

An example of this type of judgment was given to the observers using the length of lines. The observers were then warned against: 1) judging the second series on the basis of the *number* of digits it contains; 2) attempting to be consistent for the sake of being consistent; 3) using their estimate of the correctness of the digit series as the basis of the judgment.

When the instructions had been read by the observer he was encouraged to ask as many questions as he wished. The same principle was adopted in this experiment that was adopted in the experiment on visual rate: it is more important that the observers understand the instructions than that they be given the same instructions word for word.

The observers were then given 39 groups of standard and variable. This number of judgments constituted one experimental session. There were twenty experimental sessions for each observer. Each observer made, therefore, 780 judgments and memorized or tried to memorize 1,560 digit series.

The standards ranged in length from 12 digits to 5 digits. During each experimental session the observers judged 6, 7, 8, 9, 10 and 11 digits against a standard of 12; 5, 6, 7, 8, 9 and 10 digits against a standard of 11; 4, 5, 6, 7, 8 and 9 digits against a standard of 10; 4, 5, 6, 7 and 8 digits against a standard of 9; 3, 4, 5, 6 and 7 against a standard of 8; 3, 4, 5 and 6 against a standard of 7; 2, 3, 4 and 5 against a standard of 6; and 2, 3 and 4 against a standard of 5 digits.

All the standards of any one length were presented together, i.e., all the judgments against the standard of 12 were completed before going on to the next standard. The variable series judged against a given standard were presented in a random order. The order in which the groups of standards were presented was randomized with respect to the daily sessions.

The digits were presented at approximately one per second in an even tone of voice. Rest periods were allowed the observers when they requested them.

The digit series were constructed by the experimenter. No two numerals that are adjacent in the numeral series were ever adjacent in the digit series, and no regular progressions, such as 3, 5, 7, were used. 780 of such series were constructed. When they had been used once, which required 10 daily sessions, the 780 were repeated.

SECTION D. RESULTS

In the case of each standard the percentages of greater-than-one-half judgments were plotted for each of the variables. Probability coordinates were used (24). The points were fitted with a curve by eye and the median 1/2 judgment was read directly from the curve.

The data from one of the observers was fitted by three different methods, best fitting straight line (by eye), best fitting curve (by eye) and best fitting straight line by the Müller-Urban weights.

The comparison of the methods convinced the experimenter that nothing was to be gained by using the M rather than the median and that there was little to choose between the rectilinear and the curvilinear fit. The chief difference between the straight line and the curve lies at the ends of the distribution. The change in the median values was negligible. Furthermore, despite the fact that many of the plots on the probability paper showed skewness, the differences between the M and the median was not very large.

The median 1/2 judgments were separately calculated for the first ten

sessions and the second ten sessions. This was done in order to determine the effect of practice on the 1/2 judgments. In the present study only the data from the first ten sessions will be presented.

ments for the different standards, the percentage of the standards that was objectively incorrect and the percentage of the standards that was subjectively incorrect. The table also presents the M

TABLE 3

A summary of the results of the experiment on the subjective difficulty of memorizing and reproducing digit series

	Number of digits in standard	5	6	7	8	9	10	11	12
	Number on which % objectively and subjectively incorrect is based	30	40	40	50	50	60	60	60
Be.	Number judged ½			2.65	3.45	5.35	6.7	8.25	8.9
	Per cent objectively incorrect	0	0	2.5	8.0	6.0	15.0	26.6	35.0
	Per cent subjectively incorrect	0	2.5	10.0	16.0	10.0	23.0	46.5	46.8
Co.	Number judged ½	1.30	2.35	3.8	4.65	5.45	7.0	7.6	8.65
	Per cent objectively incorrect	3.0	7.5	17.5	36.0	56.0	71.5	90.0	90.0
	Per cent subjectively incorrect	0	5.0	12.5	28.0	32.0	38.4	43.0	45.0
Dr.	Number judged ½	3.75	5.35	5.25	5.60	6.25	6.55	7.15	7.4
	Per cent objectively incorrect	0	7.5	12.5	32.0	52.0	63.0	81.5	86.5
	Per cent subjectively incorrect	0	0	5.0	16.0	44.0	50.0	80.0	98.5
Ka.	Number judged ½	3.2	3.75	4.45	4.95	5.5	5.55	6.15	6.6
	Per cent objectively incorrect	3.3	2.5	0	16.0	22.0	45.0	63.0	77.0
	Per cent subjectively incorrect	0	2.5	0	4.0	4.0	25.0	40.0	54.0
Ge.	Number judged ½	1.3	3.65	3.4	4.65	5.75	6.0	7.2	7.6
	Per cent objectively incorrect	6.0	20.0	65.0	88.0	96.0	98.5	100.0	100.0
	Per cent subjectively incorrect	0	0	2.5	10.0	18.0	21.8	52.0	55.0
	Average judged ½	1.94	2.83	3.93	4.68	5.67	6.43	7.26	7.78

When the median 1/2 judgments had been determined the percentage of incorrect responses was calculated for digit series of different length. These calculations were made only for the standards and only for the data from the first ten sessions.

Then the percentage of subjectively incorrect responses was calculated for digits of different length. These calculations were also based only on the standards for the first ten sessions.

Table 3 shows the median 1/2 judgements for the different standards, of the 1/2 judgments for all the observers. The last column in the table gives the number of cases on which the objective and subjective incorrectness was based. (Since the standard of 12 occurred in combination with series of 6, 7, 8, 9, 10 and 11 digits and the standard of 5 occurred in combination with series of 2, 3 and 4 digits, it is obvious that the standard of 12 occurred more frequently than the standard of 5.)

In Figures 29-33 the median 1/2 judgments have been plotted against the

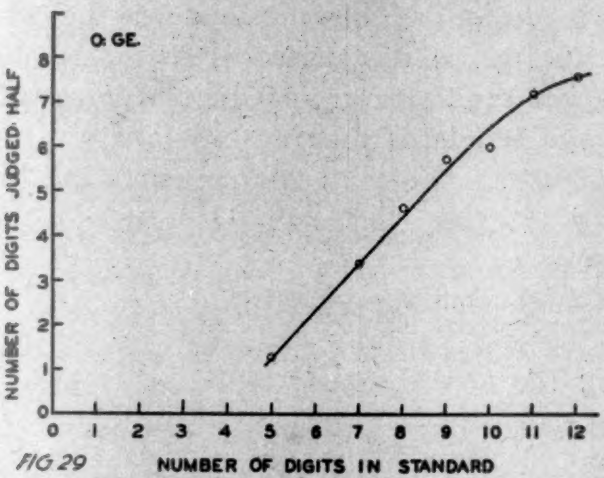

FIG 29

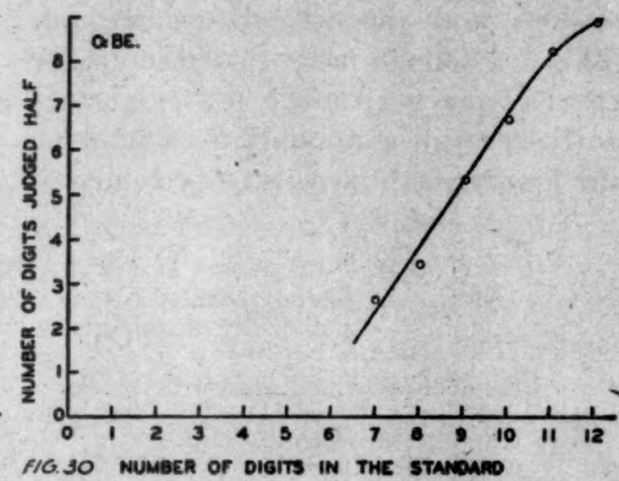

FIG. 30

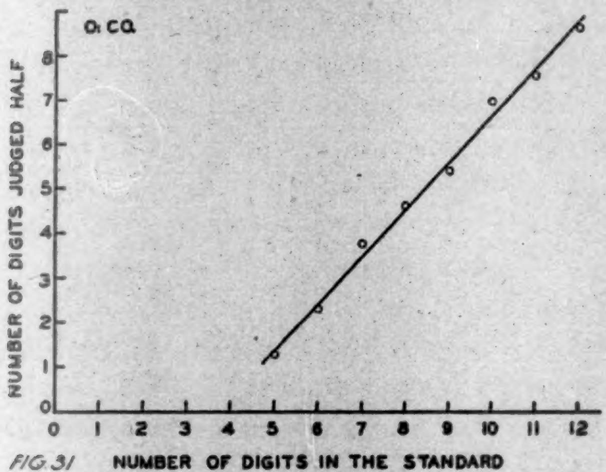

FIG 31

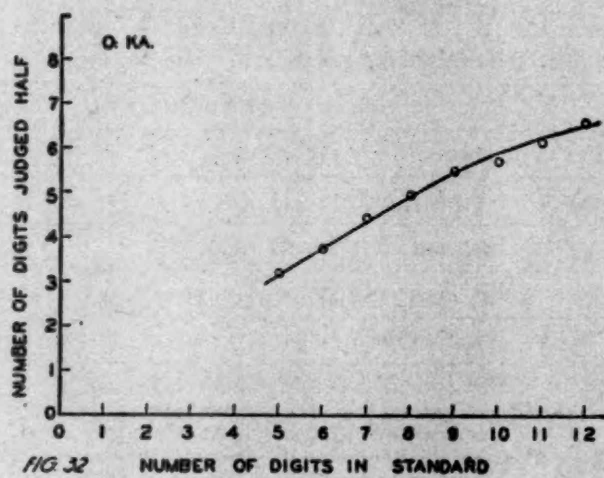

FIG 32

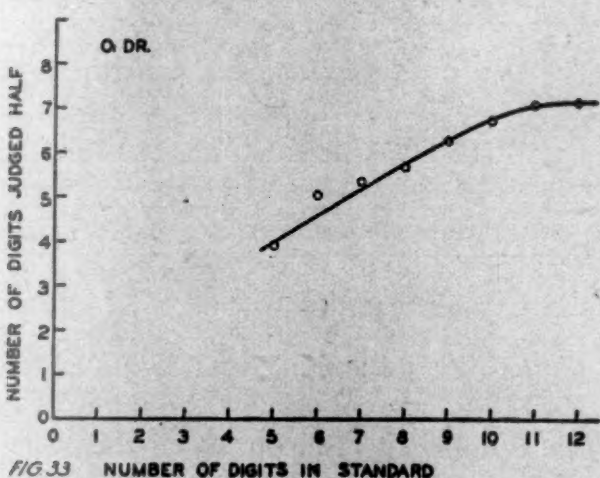

FIG 33

FIGS. 29-33. Half judgment functions for the subjective difficulty of digit series for five observers (arithmetic coordinates). The coordinates represent the actual number of digits in the series.

standards for each observer separately and the average 1/2 judgment function for all the observers is shown in Figure 34. The points have been fitted with a curve by eye. It will be noted that four of the functions have the same shape. They are approximately a straight line over the lower section of the range and become negatively accelerated near its upper end. The fifth function, that for *Co.*, is best fitted by a straight line. The similarity between the shapes of the 1/2 judgment functions is striking but even more striking is the smoothness of the plotted data. This is one of the best arguments obtainable for the reliability of a function.

From these curves the magnitude func-

tions have been constructed in the usual fashion. In all of the magnitude functions the point of origin is taken to be four digits on the abscissa and four units of subjective magnitude on the ordinate. In other words, four units of subjective magnitude were arbitrarily assigned to represent the subjective difficulty of four digits. These magnitude functions are shown in Figures 35-39, and the average function in 40.

In the same figures the percentage objectively incorrect and the percentage subjectively incorrect have been plotted and a smooth curve has been drawn through the points. The actual values of these percentages are not shown as plotted, but they may be found in Table 3.

SECTION E. DISCUSSION OF RESULTS

The first problem in the results concerns the negative acceleration of the upper end of the 1/2 judgment function for all the observers except *Co.* The interpretation of this deceleration is that after a certain point increases in the number of digits in the standard lead to less and less of an increase in the number of digits in the variable. In other words the subjective difficulty of memorizing and reproducing the series is reaching a limit. All of the functions show that an increase in the number of digits, when the absolute number of digits in the series is small, adds less subjective difficulty than an equal increase at the middle of the range. This is amply supported by the introspections of the observers. All of the observers except *Co.* reported that after a certain point was reached it was hard to tell the difference between one series and another. For example 11 digits were about as difficult as 12. This is certainly reasonable. If an observer has a span of 8 digits and never reproduced 12 correctly, it would seem that the addition of further digits is not going to add much to the subjective difficulty of the task. The calculus and high school algebra have equal subjective difficulty for the I.Q. of 50.

The one exception to this is observer *Co.*, who reported that the addition of a digit at 11 digits was noticeable. It will be noted that the 1/2 judgment function

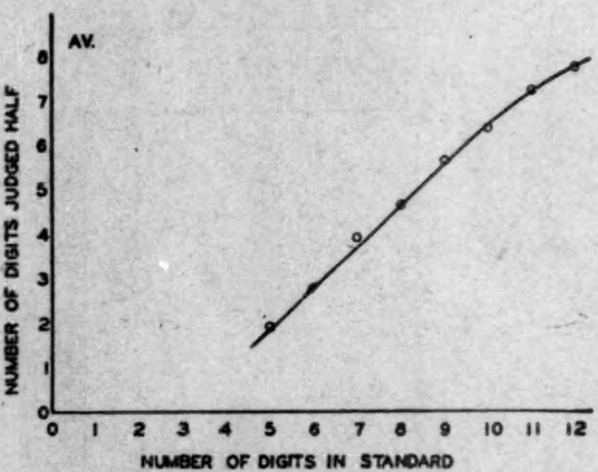

FIG. 34. The half judgment function for the subjective difficulty of digit series for five observers averaged (arithmetic coordinates). The coordinates represent the actual number of digits in the series.

for *Co.* does not show the negative acceleration shown by the other observers. The experimenter believes that *Co.'s* 1/2 judgment function would have the same shape if longer digit series had been given.

The next question concerns the establishment of zero subjective difficulty. If the magnitude functions for *Co.*, *Be.*, *Ka.* and *Ge.* were extrapolated to the ordinate, it can be readily seen that the result would be to assign a certain amount of subjective difficulty to a series of zero digits. In other words memorizing and producing no digits at all would have a certain difficulty for these ob-

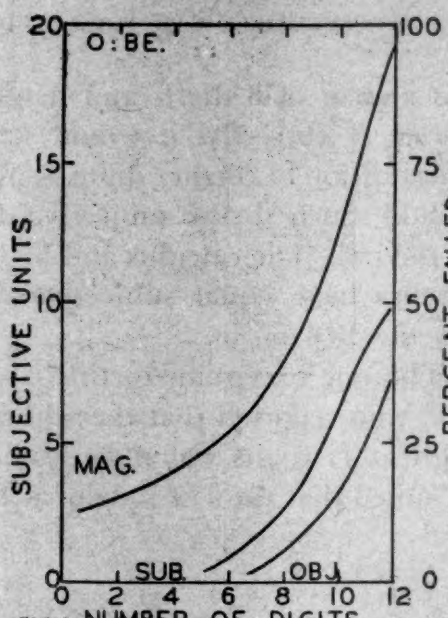

FIG.35

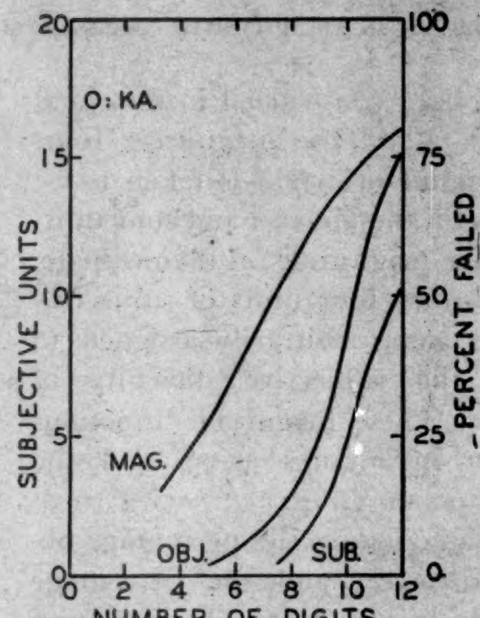

FIG.36

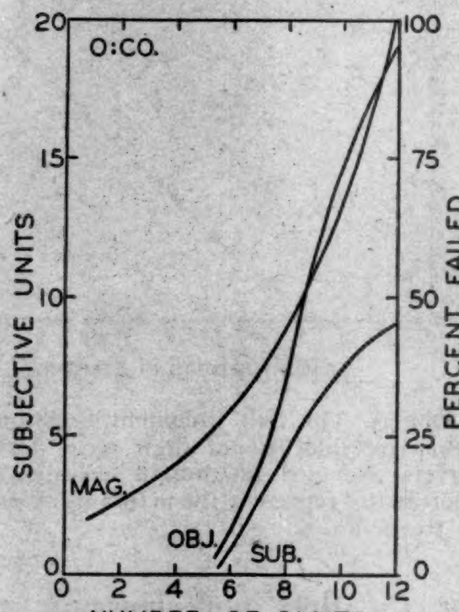

FIG.37

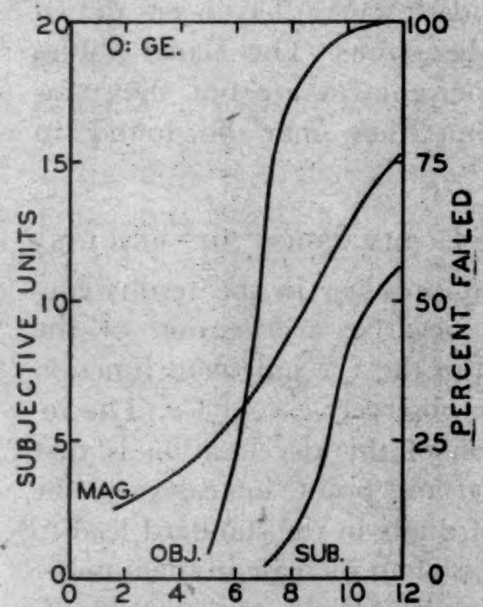

FIG.38

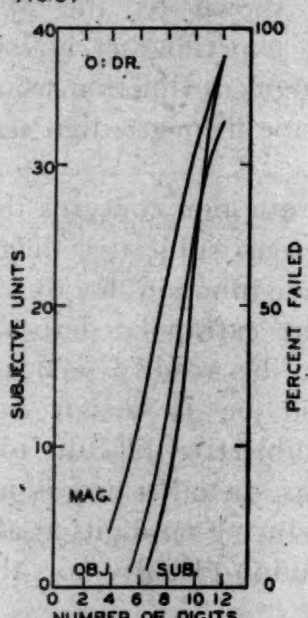

FIG.39

FIGS. 35-39. The magnitude functions for the subjective difficulty of digit series for five observers. The functions for the per cent objectively incorrect and the subjectivity incorrect are also shown. The abscissa represents the number of digits in a given series, and the ordinate is subjective difficulty (left hand side) and percentage incorrect (failed) (right hand side). The magnitude function is labelled MAG and the subjectively incorrect and the objectively incorrect are labelled SUB and OBJ respectively. The coordinates are arithmetic.

servers. The obvious extrapolation of *Dr.'s* magnitude function would be to the abscissa in the neighborhood of 3 digits. This would be interpreted that a series of 3 digits had zero subjective magnitude for this observer.

But all of the observers report either that 2 or 3 digits have no subjective difficulty or that their subjective difficulty is extremely small. All of the observers were agreed that 1 digit had no subjective difficulty at all. Yet the "obvious" extrapolation of the magnitude functions of observers *Be., Ge., Ka.* and *Co.* is to the ordinate.

The answer to this problem is fairly obvious. First, in order to obtain the median 1/2 judgment for any standard it is necessary to be able to plot at least two points on the probability coordinates. These two points must be greater than 0% and less than 100%. Suppose that two points are obtained; if they lie far from 50% and on the same side of 50% an accurate determination of the median by extrapolation is difficult and unreliable. Second, the observers reach a limit in subjective difficulty at both ends of the curve. At the upper limit, at maximum subjective difficulty, it is always possible to present the observer with variables that can be judged both greater and less than 1/2 as difficult. In other words it is always possible to obtain several points on both sides of the 50% point on the probability paper. However at the lower end of the range, at minimum subjective difficulty, it is not possible to do this. A point is reached in the neighborhood of 4 or 5 digits where the variables would always be judged as less than 1/2 as difficult. If the variable is never judged to be greater than 1/2 as difficult as the standard, or equal to 1/2 the difficulty of the standard, then the percentage of greater judgments is 0. No points can be obtained by which the median 1/2 judgment can be found. Furthermore, even if an occasional greater or equal judgment is given for two of the variables, the points will lie far from the 50% points on the probability paper and extrapolation is either impossible or extremely hazardous. This means, then, that points near zero subjective magnitude will be impossible to obtain. In short, the true direction of extrapolation of the magnitude function plots is experimentally indeterminable

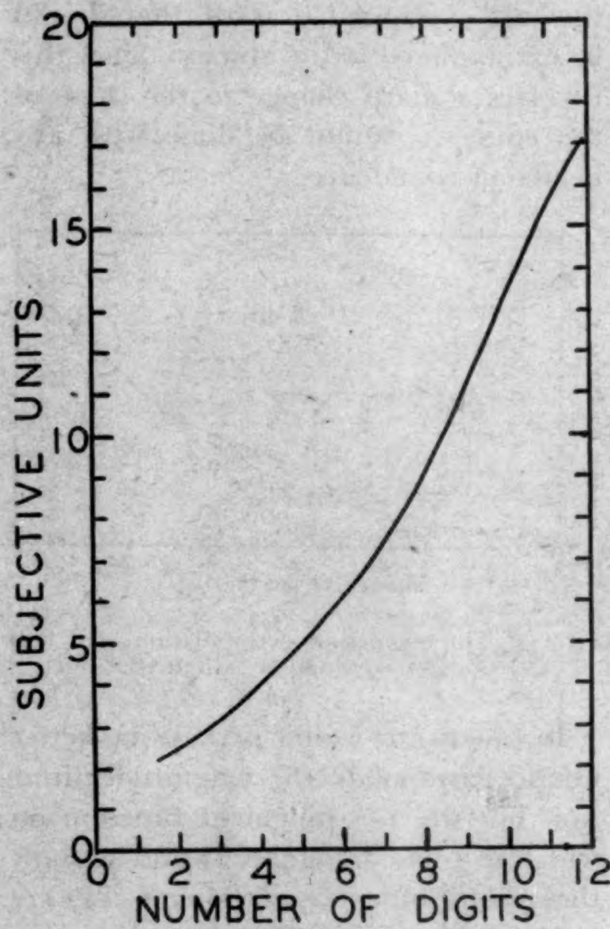

FIG. 40. The magnitude function for the subjective difficulty of digit series for the average of five observers (arithmetic coordinates). This function was constructed from Figure 34. The abscissa represents the number of digits in the series and the ordinate represents units of subjective difficulty.

by this method, because the true shape of the 1/2 judgment functions cannot be determined at or near zero subjective difficulty. However, reasoning from the introspective evidence and from common sense, it is readily seen that the "obvious" extrapolation of the magnitude functions to the ordinate is not correct. If they are extrapolated at all, they should be extrapolated to the abscissa. Since this involves a sharp change in the slope of the curve it cannot be done with any degree of confidence.

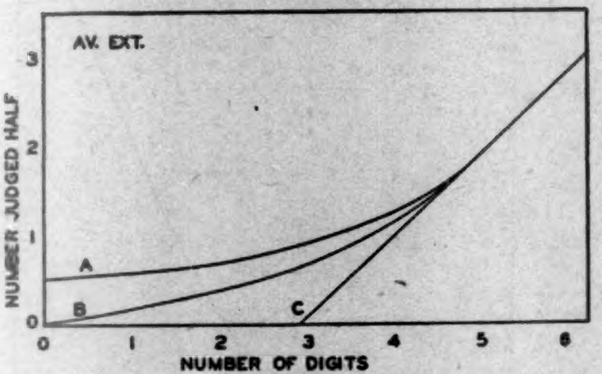

FIG. 41. Three types of extrapolation of a half judgment function (see text).

In practice it would perhaps be better not to extrapolate the magnitude function but the 1/2 judgment function on which it is based. Figures 41 and 42 show the results of extrapolation of the 1/2 judgment function. Figure 41 represents a curve for some hypothetical data. The curve has been extrapolated from the point X in the directions represented by A, B and C. Figure 42 shows the resulting effect on the magnitude function. Extrapolation of the 1/2 judgment function in the direction A results in the magnitude function A in Figure 42; likewise extrapolation of the 1/2 judgment function in the directions B and C results in the magnitude functions B and C respectively.

It is clear, then, that the obtained data are inadequate for the determination of zero subjective difficulty. However, one thing seems certain from the introspective evidence: zero subjective difficulty is not associated with less than 1 digit.

An examination of the objectively and subjectively incorrect judgments leads to some rather interesting considerations. The order of the systems in the magnitude functions and in the objectively incorrect and subjectively incorrect functions are the same. They are magnitudes of the same kind. The common factor is obviously the number of digits in the series.

Furthermore all of the objectively incorrect and subjectively incorrect functions have the same shape. (*Be.'s* objectively incorrect curve has not become negatively accelerated at 12. *Be.* has a digit span of about 14 digits, but it can be assumed with maximum confidence that he will eventually reach a series where he will fail more often than he succeeds!)

The similarity between the shape of the magnitude functions and the other two functions cannot be readily explained on the basis of some factor common to all three. Of course there is a common factor, the actual number of digits, but this merely establishes the fact that the systems have the same order. The objectively incorrect and subjectively incorrect functions are typical psychometric functions. The magnitude function has an upper asymptote, but the deceleration at the top of the magnitude function is due, it has been said, to the approach of an upper limit of difficulty beyond which no increase in number of digits adds to the subjective difficulty of the task.

Some other interesting facts can be gleaned from a study of the functions. The observer *Dr.* was extremely em-

barrassed about serving as a subject. Although his span is well above average, he thought that it was low and did not want the experimenter, who was a friend, to find out how low it was. He found the task extremely difficult and worked very hard to do well. All of these facts are reflected in the functions. His magnitude function rises very steeply. He finds the task of memorizing the digits extremely difficult but because of his superhuman labor he became confident about the shorter digits. His confidence is greater than his performance for the lower regions of the curve. Only at the upper end of the curve did his feelings of inferiority overcome his effort. He was sure that he was not getting any of the series correct, when in actuality he was.

Three of the five observers were more confident than their objective performance warranted.

A mere examination of the objective and subjective digit spans of these observers would not give all the information obtained by actually plotting out the functions. A good example of this is found in *Dr.'s* functions discussed above. The crossing of the subjective and objective functions at about 10.3 digits is, as was seen in the discussion, one of the interesting facts in this observer's performance. Likewise the increasing spread between the objective and subjective functions of *Ka.* would not be seen by merely comparing two points on the curve. *Ka.* was more confident than her performance warranted and the confidence seemed to be not too greatly affected by objective incorrectness. Throughout the experimental sessions she seemed to be quite self-assured.

SECTION F. SUMMARY AND CONCLUSIONS

1) The problem was to test the possibility of constructing an equal unit scale for the subjective difficulty of memorizing and reproducing digit series.

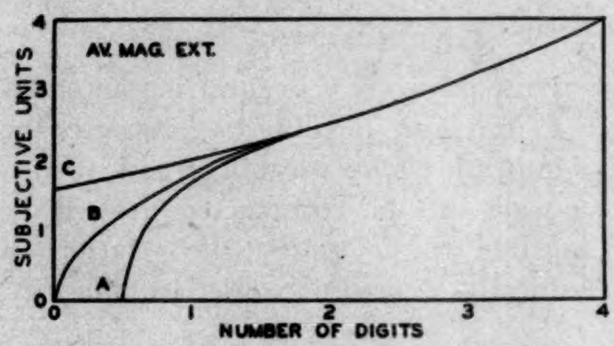

FIG. 42. The three types of magnitude function resulting from the extrapolations shown in Figure 41 (see text).

2) The scales were successfully constructed for each of five observers.

3) The 1/2 judgment functions are the same shape for 4 of the 5 observers and the points for any one observer formed a smooth, continuous function. For this reason the experimenter considers the results reliable.

4) The deceleration of the magnitude functions was due to the approach of an upper limit of subjective difficulty.

5) The introspections of the observers indicated that there were many complex determinants of subjective difficulty.

6) The analysis of the subjectively incorrect and objectively incorrect functions showed some individual characteristics paralleled by the observers' behavior during the experiment.

7) The fact that this impalpable characteristic has been successfully scaled by an equal unit scale indicates that this type of material is open to empirical investigation by other than purely statistical techniques.

PART V

EXPERIMENTAL: THE SCALING OF THE SUBJECTIVE DIFFICULTY OF WORDS IN A MULTIPLE CHOICE VOCABULARY TEST

SECTION A. THE PROBLEM

THE PROBLEM is to construct an equal unit scale of subjective difficulty for a multiple choice vocabulary test. If such a scale can be constructed, it will be possible to determine the relation of subjective difficulty to objective difficulty.

SECTION B. DISCUSSION OF THE PROBLEM

It will be remembered that the subjective difficulty of a vocabulary test was cited as an example of a characteristic that either has no stimulus correlate at all or that thas one that is exceedingly complex. The scaling of such a characteristic was discussed in Part I. The conclusion drawn from that discussion was that no other measurable magnitude is necessary for the construction of a subjective scale. All that is necessary is to have identifiable systems.

There are several ways in which vocabulary items may be identified; the first and most obvious is by the word itself, since the words to be defined in different items are different; second, if the words are the same but have different multiple choice words, the item can be identified by the multiple choice words; third, it can be identified by its objective difficulty. If the items are identified by the third means all items with the same objective difficulty are in principle indistinguishable, i.e., they are indistinguishable so far as objective difficulty is concerned. The first necessity in constructing the equal unit scale is to establish an ordinal scale. This was not done either in the case of speed or in the case of the digit series; from previous knowledge and experience there is maximum certainty that subjective rate increases as a function of physical rate and that the subjective difficulty of memorizing and producing the digit series increases as a function of the number of the digits in the series. In short, the order is the same. The magnitude functions prove that this is true; if it were not true the function would either change slope and descend at some point or there would be reversals in the percentage judged greater than 1/2 in the raw data.

In the present case it was assumed that subjective difficulty would be positively correlated with percentage of the total population passing the item, so no ordinal scale was constructed. This assumption will be discussed later. As in the case of visual rate and the subjective difficulty of the digit series a ready made ordinal scale was at hand.

The vocabulary items were obtained from the I.E.R. Inventory of intellectual tasks and their difficulty (53). These items have been scaled in objective difficulty by a statistical scaling technique described by Thorndike (47). The units in which the items are scaled are supposedly "equal" and supposedly measured from an "absolute zero." In Part I there was a discussion of the "equality" obtained by the statistical techniques and no more need be said of it here. In plotting subjective difficulty against Thorndike's units, the question arises, how should Thorndike's units be spaced

on the abscissa? There is no rule for spacing his units unless his scale is considered as a B-magnitude with a postulated relation between "difficulty" and the units. This matter was also discussed in Part I.

The difficulty of the items in Thorndike's units ranged from 220 to 442 or, in other words, if one accepts Thorndike's units as equal and his zero as absolute, the most difficult item was about "twice" as difficult as the easiest. These units, based on percentage passing, are extremely fine. There seemed to be no detectable change in subjective difficulty over a five point range. So to simplify the collection of the material for the experiment, the total range of Thorndike's units was divided into 43 new large units, each unit consisting of a range of approximately 5 of Thorndike's units. The new unit 1 contained Thorndike's words of difficulty 220-224; 2 contained 225-229; 3 contained 230-234; 4 contained 235-239, etc. The chief exception to this was the new unit number 43 which contained 431-442. Several of the new units contained only four of Thorndike's units.

The vocabulary items of 220-300 of Thorndike's units are multiple choice picture vocabulary items. These items are numbers 1 to 16 in the new units. Values in the succeeding discussion will be stated in the new units. A summary of the criticisms leveled against the scaling of this test by Thorndike will be found in Gansl (17). Despite the faults of the scale it is highly probable that what errors there are would not be of sufficient magnitude to affect the relation between the percentage passing and subjective difficulty.

It is now necessary to return to the discussion of the identification marks of the systems. In the present case the identification marks are the new units 1-43. (Within any one of these units the separate tests are, of course, "identified" by the word to be defined.) These units will also be used on the abscissa of the magnitude function. An objection may be raised that these units are not equal in the sense that Thorndike's original units were equal because they do not contain an equal number of Thorndike's units. The answer is, first, most of them do contain an equal number of Thorndike's units; second, the error introduced is certainly no larger than that necessarily introduced by the random selection of the items within any one of the units. The one exception is new unit 43 which contains ten of Thorndike's units. To compensate for this difference 43 has been plotted twice the usual distance from the unit next below it.

SECTION C. PROCEDURE

The procedure was similar to that used in the determination of the median 1/2 judgment for the digit series. There were ten standards and each standard was compared with a series of variables. The task of the observer was to give a judgment of "greater than 1/2 as difficult," "1/2 as difficult" or "less than 1/2 as difficult." At some levels there were not enough items to pair the standard with more than one of each of the values of the variable. Table 4 gives the values of the standards and the variables with which they were compared together with the number of comparisons for each of the values of the variable.

The items were printed singly on separate sheets of paper and presented in pairs to the observer. He was instructed to do the first item (the standard) and then to do the second item (the variable) and make his judgment. The instructions were similar to the instructions for the digit series; the only real difference

TABLE 4

Showing the values of the standards, the variables with which they were compared and the number of judgments made for each comparison by each observer

Standard	Variable	N
43	40	1
43	34	1
43	31	1
43	28	1
43	25	1
42	39	1
42	36	1
42	33	1
42	27	1
42	24	1
41	38	1
41	35	1
41	32	1
41	26	1
41	23	1
39	37	1
39	32	1
39	29	1
39	25	1
39	22	1
37	35	2
37	31	2
37	28	2
37	25	2
37	22	2
34	32	3
34	29	3
34	26	3
34	22	3
34	16	3
31	28	3
31	25	3
31	22	3
31	12	3
25	22	2
25	16	2
25	8	2
25	5	2
15	13	1
15	9	1
15	6	1
15	3	1
6	4	1
6	3	1
6	2	1
6	1	1

was due to the nature of the material. The pairs of words were presented in a random order to the observers. The pairing of standard and variable for any one comparison was randomized, i.e., the comparison of 43 and 40 did not mean that the identical items were paired for every observer but only that *an* item with the value 43 was compared with *an* item of 40.

There were fifteen observers.

SECTION D. RESULTS

The number of "greater than 1/2" judgments made by each of the observers for each of the variables for any given standard was added. The sums were converted into percentages and plotted on probability paper, and the median 1/2 judgment obtained in the way outlined for the digit series.

Table 5 shows the median 1/2 judgment for all the observers together for each of the standards.

The data have been plotted and a curve fitted to the points by eye (Fig. 43). From this curve a magnitude function has been plotted in the usual manner. The origin is at 10 objective units on the abscissa and 10 subjective units on the ordinate (Fig. 44).

SECTION E. DISCUSSION OF RESULTS

The first thing to note about the curve in Figure 43 is the reversal at its upper end. It will be remembered that it was assumed that subjective difficulty would be correlated positively with the percentage of the general population passing the items. This is true for most of the range of difficulty but ceases to be true at the upper end of the curve. In other words, the assumption that the two functions have the same order was not correct for the entire range. There is some introspective evidence that might explain this. The upper levels of difficulty contain multiple choice words

TABLE 5

Showing the median ½ judgment for each of the standards

Standard	43	42	41	39	37	34	31
Median ½ judgment	34.8	36.5	35.0	35.0	22.8	14.0	10.0

which have a certain confusion-value for the observer. The word to be defined is so difficult that the observer seizes on the most likely (but often incorrect) word. He bases his choice on the apparent similarity of the stems or on some other false criterion. But the net effect of this is not to make the task subjectively more difficult. The reversal may not be completely explained on this basis, but it is at least partially explained.

The assumption that the subjective difficulty of the items and the objective difficulty would have the same order seemed a safe assumption. The results of Farmer (12), of Hertzman (25) and of the experimenter's own digit span experiment had led him to believe that the assumption was probably valid.

A more correct procedure would be first to order the systems empirically; second, to group together all those systems that were judged to be subjectively equal; third, to scale these groups of systems in a manner similar to that used in the present experiment.

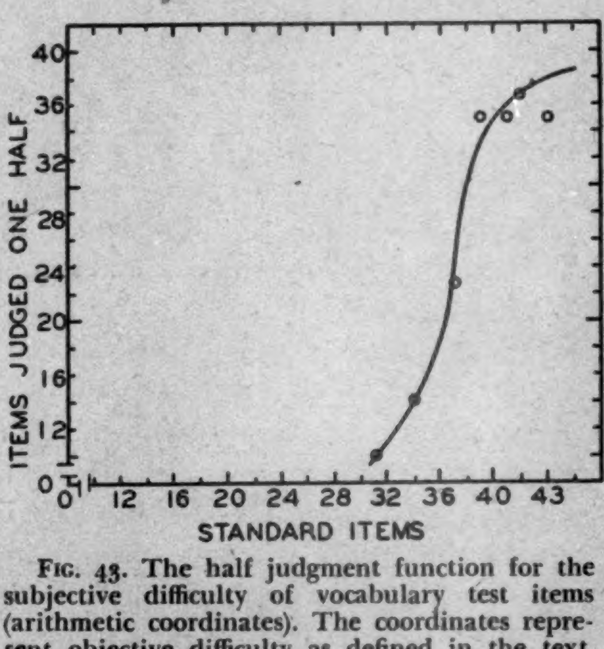

FIG. 43. The half judgment function for the subjective difficulty of vocabulary test items (arithmetic coordinates). The coordinates represent objective difficulty as defined in the text.

One technical difficulty has arisen in the construction of the magnitude function. The 1/2 judgment function has a

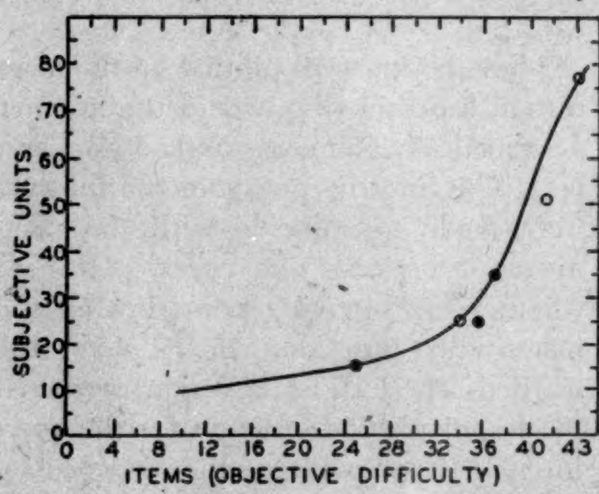

FIG. 44. The magnitude function for the subjective difficulty of vocabulary test items (arithmetic coordinates). The abscissa represents objective difficulty and the ordinate represents subjective difficulty. The circles are explained in the text.

very steep slope which means that few points can be obtained for the construction of the magnitude function. The result is that it is difficult to fit a curve to points of the magnitude function. The only way that this can be done is by trial and error, checking the curve for internal consistency as described in the discussion of discontinuous functions.

The obtained magnitude function was checked by the method of equal appearing intervals.[33] The procedure was similar to that adopted for the main part of the experiment. In this case, however, the observers were given three items and were told to do all of the items and then

[33] This experiment was carried out by W. R. Brunner, Columbia College.

to judge whether the interval of subjective difficulty between the first and second item was "greater than" "equal to" or "less than" the interval in difficulty between the second and third items. The results were handled in the same manner as in the main experiment. Two intervals were used: 43-34 and 37-25. The interval 47-41.6 was judged to be equal to 41.6-34 and the interval 37-35.8 was judged to be equal to the interval 35.8-25.

These points were plotted on the magnitude function (Fig. 42) in the manner described by Stevens and Volkmann (42). The limiting points of the interval were made to coincide with the magnitude function. The check is to see whether the bisecting point lies on the magnitude function. If it does the methods check. The bisection predicted by the magnitude function for the 47-34 interval is 39.8 whereas the obtained bisection is 41.6; the predicted bisection of the 37-25 interval is 34 and the obtained bisection is 35.8. The agreement is not perfect but the error is not great.

SECTION F. SUMMARY AND CONCLUSIONS

1) The problem was to construct an equal unit scale for the subjective difficulty of multiple choice vocabulary items.

2) The scale was constructed, but it possesses certain shortcomings. The assumption that the order of the items with respect to subjective difficulty was the same as the order with respect to percentage passing broke down at the upper end of the function. The correct way of constructing the function without making this assumption is described.

3) The magnitude function was checked by the method of equal appearing intervals. The results of this operation did not agree perfectly with the magnitude function but the margin of error was not great.

General Summary

The purpose of this study was to examine the possibility of measuring all those discriminable characteristics that exist in discriminable degrees.

The logical criteria for measurement were critically examined and the psychological scaling methods were analyzed in the light of these criteria.

The conclusion was drawn that none of the attempts at measurement, used so far by psychologists, meet the necessary criteria for fundamental measurement. It was argued that the equal unit scale meets more of the necessary criteria than any other set of operations. The further conclusion was drawn that there is no requirement for measurement of any type that psychology in principle cannot meet, since the criteria of physical juxtaposition is not a requisite for addition.

Equal unit scales were constructed for three discriminable characteristics that differed in important respects from each other and that presented special theoretical and practical problems. It is concluded that it is possible to construct equal unit scales, i.e., scales that meet the logical requirements for order and equality of units for discriminable characteristics that are both impalpable and without a known correlated stimulus correlate.

BIBLIOGRAPHY

1. BARTLETT, R. J. Measurement in psychology. *Advancement of Sci.*, 1940, No. 3, 422-441.
2. BINET, A., & SIMON, TH. Méthodes nouvelles poure faire le diagnostic differentiel des anormaux de l'intelligence *Année psychol.*, 1905, *11*, 191-244.
3. BORING, E. G. A history of experimental psychology. New York: D. Appleton-Century, 1929.
4. BORING, E. G. The physical dimensions of consciousness. New York: Century, 1933.
5. CAMPBELL, N. R. Physics. The elements. Cambridge: The University Press, 1920.
6. CAMPBELL, N. R. An account of the principles of measurement and calculation. London: Longmans, Green, 1928.
7. CAMPBELL, N. R. Symposium: Measurement and its importance for philosophy. *Proc. Aristotelian Soc. Suppl.* Volume 17. London: Harrison & Sons, 1938.
8. CAROTHERS, F. E. Psychological examinations of college students. *Arch. Psychol., N.Y.*, 1921, *6*, No. 46.
9. COHEN, M. R. & NAGEL, E. An introduction to logic and scientific method. New York: Harcourt, Brace, 1934.
10. DUNLAP, K. Difference-sensibility for rate of discrete impressions. *Psychol. Rev.*, 1912, *19*, 32-59.
11. DUNLAP, K. The shortest perceptible time-interval between two flashes of light. *Psychol. Rev.*, 1915, *22*, 226-250.
12. FARMER, E. Concerning subjective judgment of difficulty. *Brit. J. Psychol.*, 1938, *33*, 100-111.
13. FERGUSON, A., MYERS, C. S., BARTLETT, R. J., *et al.* Quantitative estimates of sensory events. *Rep. Brit. Assoc. Advancement Sci.*, 1938, No. 108.
14. FERGUSON, A., MYERS, C. S., BARTLETT, R. J., *et al.* Quantitative estimates of sensory events. Final report. *Advancement of Sci.*, 1940, No. 2, 331-349.
15. FISHER, R. A., & YATES, F. Statistical tables for biological, agricultural and medical research. London: Oliver & Boyd, 1938.
16. GAGE, F. H. An experimental investigation of the measurability of auditory sensation. *Proc. roy. Soc.*, 1934, *116B*, 103-122.
17. GANSL, I. Vocabulary: its measurement and growth. *Arch. Psychol., N.Y.*, 1939, *33*, No. 236.
18. GARRETT, H. E. The relation of tests of memory and learning to each other and to general intelligence in a highly selected adult group. *J. educ. Psychol.*, 1928, *19*, 601-613.
19. GARRETT, H. E., & SCHNECK, M. R. Psychological tests, methods and results. New York: Harper & Brothers, 1933.
20. GIBSON, J. J. Adaptation and negative after-effect. *Psychol. Rev.*, 1937, *44*, 222-244.
21. GUILFORD, J. P. Psychometric methods. New York: McGraw-Hill, 1936.
22. GUILFORD, J. P. The psychophysics of mental test difficulty. *Psychometrica*, 1937, *2*, 139-150.
23. HARTMANN, G. W. The relative influence of visual and auditory factors in spelling ability. *J. educ. Psychol.*, 1931, *22*, 691-699.
24. HAZEN, A. Storage to be provided in impounding reservoirs. *Trans. Amer. Soc. Con. Eng.*, 1914, *77*, 1539-1669.
25. HERTZMAN, M. Confidence ratings as an index of difficulty. *J. exp. Psychol.*, 1937, *21*, 113-119.
26. HEVNER, K. An empirical study of three psychophysical methods. *J. gen. Psychol.*, 1930, *4*, 191-212.
27. JAMES, W. The principles of psychology. Vol. I. New York, Henry Holt, 1890.
28. JOHNSON, H. M. Pseudo-mathematics in the mental and social sciences. *Amer. J. Psychol.*, 1936, *48*, 342-351.
29. KUHLMANN, F. Tests of mental development. Minneapolis: Educational Publishers, 1939.
30. MARTIN, P. R., & FERNBERGER, S. W. Improvement in memory span. *Amer. J. Psychol.*, 1929, *41*, 91-94.
31. MCGREGER, D. Scientific measurement and psychology. *Psychol. Rev.*, 1935, *42*, 246-266.
32. NEWMAN, E. B., VOLKMANN, J., & STEVENS, S. S. On the method of bisection and its relation to a loudness scale. *Amer. J. Psychol.*, 1937, *49*, 134-137.
33. PEATMAN, J. G. On the meaning of a test score in psychological measurement. *Amer. J. Orthopsychiat.*, 1939, *9*, 1939.
34. SCATES, D. E. The essential conditions of measurement. *Psychometrika*, 1937, *2*, 27-34.
35. SCATES, D. E. How science measures. *J. exp. Educ.*, 1937, *5*, 296-312.
36. SMITH, B. O. Logical aspects of educational measurement. New York: Columbia University Press, 1938.
37. SPENCER, L. J. Abrasives. *Ency. Brit.* 14th Ed., Vol. 1, 1929, 61-62.
38. STEVENS, S. S. A scale for the measurement of a psychological magnitude, loudness. *Psychol. Rev.*, 1936, *43*, 405-416.

39. STEVENS, S. S., VOLKMANN, J., & NEWMAN, E. B. A scale for the measurement of the psychological magnitude pitch. *J. acoust. Soc. Amer.*, 1937, *8*, 185-190.

40. STEVENS, S. S. On the problem of scales for the measurement of psychological magnitudes. *J. Unified Sci.*, 1939, *9*, 94-99.

41. STEVENS, S. S., & DAVIS, H. Hearing: its psychology and physiology. New York: John Wiley, 1938.

42. STEVENS, S. S., & VOLKMANN, J. The relation of pitch to frequency: a revised scale. *Amer. J. Psychol.*, 1940, *53*, 329-353.

43. TABACK, M. The magnitude function for the intensity of lifted weight. Columbia Univ. Psychol. Lab., 1940. Unpublished.

44. TAVES, E. H. Two mechanisms for the perception of visual numerousness. *Arch. Psychol., N.Y.*, 1941, *37*, No. 265.

45. TERMAN, L. M. The measurement of intelligence. New York: Houghton Mifflin, 1916.

46. TERMAN, L. M., & MERRILL, M. A. Measuring intelligence. New York: Houghton Mifflin, 1937.

47. THORNDIKE, E. L., *et al.* The measurement of intelligence. New York: Teach. Coll., Columbia University, 1926.

48. THURSTONE, L. L., & CHAVE, E. J. The measurement of attitude. Chicago: The University of Chicago Press, 1929.

49. TINKER, M. A. *in* Boring, E. G., Langfeld, H. S., & Weld, H. P. Psychology: a factual textbook. New York: John Wiley, 1935.

50. VOLKMANN, J. A timing relay for psychophysical experiments. *Amer. J. Psychol.*, 1938, *51*, 731-734.

51. WILLOUGHBY, R. R. The meaning of "difficulty." *Amer. J. Psychol.*, 1936, *49*, 164-167.

52. WOODWORTH, R. S. Experimental psychology. New York: Henry Holt, 1938.

53. The Division of Psychology, Institute of Educational Research, Teachers College, Columbia University. The I. E. R. inventory of intellectual tasks and their difficulty. New York: Bureau of Publications, Teachers College, Columbia University, 1926.